Author:
Sandhu, Sohaib

Title:
Exploring diagnostic listening assessment in the classroom

A Saudi academic context

Exploring diagnostic listening assessment in the classroom - A Saudi academic context.

Sohaib Sandhu

Graduate School of Education

University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Education in The Faculty of Social Sciences and Law, Graduate School of Education

October 2018

Word count: 52209

Abstract

If there is an aspect of language testing and assessment that has been under-researched and needs urgent attention, it is diagnostic testing (Alderson, 2005). It has found brief mention in the language testing literature, but there have even been instances where there has been no mention whatsoever in books claiming to be about language testing. Research has slowly but gradually begun to come through, though in reality there is still much to be done. Seen as an area of testing that focusses on finding learner strengths and weaknesses, with the potential to aid teachers in providing much needed information to guide them towards appropriate remedial instruction, there is still very limited teacher involvement in the process. Most of the diagnostic research to date (Alderson, 2010) describes the use of sophisticated diagnostic models to extract granular information, very often from tests that were never intended to be diagnostic in nature, and by researchers who may never have ever entered the classroom (Davidson, 2010). Additionally, large scale online and computer-based diagnostic systems have been developed by highly trained testing and assessment researchers for university contexts which ultimately provide diagnostic reports for use by teachers. Most of the handful of classroom based examples have required the intervention of researchers; however, rarely has there been a demonstration of a sole teacher being at the centre of the diagnostic process even in the classroom.

This research is therefore an attempt to assess the feasibility of embedding the diagnostic process into the daily task of teaching, where diagnostic tool development and implementation is teacher-led, and where diagnosis is not something that happens infrequently, but in fact is a daily activity in the classroom. For this research, the skill in focus is listening, often seen as a complex receptive skill that is not fully understood by especially classroom teachers. A listening diagnostic self-assessment tool was developed and integrated into a traditional listening lesson, the intention being to ascertain problems listeners had with each listening passage. Using an action research approach, data was collected twice weekly, over a period of five weeks, ending with interviews of the research participants. The resulting data produced for each listening lesson included a class profile, as well as individual student profiles which encompassed data for all of the listening lessons, indicating that it is feasible for teachers to collect diagnostic data in the classroom.

Overall, the results indicate that it is possible to incorporate diagnostic assessment, and specifically diagnostic listening assessment into a routine classroom lesson, which provides information that, would normally not be available in a classroom context, and which a teacher could find useful in gaining insights into learner strengths and weaknesses, allowing teachers to consider relevant remedial steps to help learners. The impact of the research and its possible applications, along with possible areas of further research are suggested.

Acknowledgements

There are many people at Bristol who impacted on my journey to the completion of this doctorate. These include in the beginning Dr.Richard Kiely, who was a source of encouragement and a reason for joining Bristol and who left a year or so later. Dr. Francis Giampapa, my academic supervisor who ensured that I pulled my socks up when needed, especially in the beginning when I was still developing my research skills. Later on, despite my reluctance, she also convinced me to take a break from studies after my mother and younger brother passed away. Thank you.

Dr. Talia Isaacs, who was my initial supervisor at dissertation stage, helped me to develop my research ideas and believed in my ability to do the doctorate. Her enthusiasm was infectious and resulted in a positive start. After her departure, I am very grateful to Dr. Helen Woodfield who took over. Her experience, patience and focussed feedback helped me to successfully complete my dissertation. With only three months left to completion, Helen came in not knowing what my research was about. However, having her as my supervisor reassured me that I was in safe hands. Thank you so much, I really do appreciate your support. Thank you also to the other lecturers who taught the various EdD modules. I learnt something on each course. Travelling from Saudi Arabia, doing the pre-reading whilst on the plane, attending a three day module, and then quickly flying back, made it all exciting. Thank you also to the administrative staff at the GSOE who were always very professional in their dealings over the years.

To my friend, brother, former colleague and Bristol EdD graduate, Muhibullah Abdur-Rahman, thank you for all of those conversations, and those brainstorming sessions we had. Having someone to talk to about my research, brought sanity to the whole process.

I owe the greatest thanks and gratitude to my wife Asma, and daughters, Madiha, Nabiha and Haadiya. My wife who was a pillar of strength, was always there when I needed some extra words of encouragement and support, despite her, at times, own ill-health. To the kids, I know there were times we could not go out simply because ‘dad’ (Abaji) was busy. Yours and your mother’s patience was exceptional.

Thank you Allah (God)

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original, except where indicated by special reference in the text, and no part of the dissertation has been submitted for any other academic award. Any views expressed in the dissertation are those of the author.

Contents

List of Abbreviations

AWE	Automated Writing Evaluation
CAEL	Canadian Academic English Language Test
CBA	Computer-Based Assessment
CDA	Cognitive Diagnostic Assessment
CDM	Cognitive Diagnostic Model
CDPM	Cognitively Diagnostic Psychometric Model
CEFR	Common European Framework
COMP-DA	Computerised Dynamic Assessment
DELA	Diagnostic English Language Assessment
DELNA	Diagnostic English Language Needs Assessment
DELTA	Diagnostic English Language Tracking Assessment
DIALANG	
DIALUKI	Diagnosing reading and writing in a second or foreign language
EDD	Empirically-derived Description-based Diagnostic-checklist
EDiT	English Diagnostic Test of Grammar
GESE	Trinity College Graded Speaking Exam
HDF	Holistic Diagnostic Feedback
HSK	Chinese Proficiency Test
IADE	Intelligent Academic Discourse Evaluator
MELAB	Michigan English Language Assessment Battery
OEAS	Online English Assessment System
PTE	Pearson Test of English
RUM	Reparamaterised Unified Model
SMW	Select Missing Word
SRST	Story Retelling Speaking Test
WM	Working Memory

List of Tables

List of figures

Chapter 1 MOTIVATION AND PURPOSE

1.1 Overview

In this chapter I briefly explain what motivated me to choose this particular area of research, followed by the research background that acts as the take-off point and justification for this research undertaking. The research questions, a synopsis of the literature and research design follow. Finally, the research timeline and a chapter brief is also presented.

1.2 Overall Research Objective

When reading the literature on diagnostic assessment that relates specifically to classroom contexts (Jang, 2009, 2013; Hirai & Koizumi, 2013; Fox & Hartwick, 2011), it seems that teachers play the role of only users or receivers of diagnostic tools, rather than being the developers. This possibly reflects the dearth of literature on classroom language assessment generally (Jin, 2010) as well as the perceived lack of assessment literacy amongst teachers (Fulcher, 2012). As a practitioner in the classroom, I wanted to conduct teacher-led research, in order to demonstrate a more teacher classroom-oriented perspective on how assessment could be implemented. Much is also written about the lack of research in diagnostic assessment (Alderson, 2005; Alderson, 2010; Lee, 2015; Mansouri, 2017), and the often cited view that it has the potential to make a real difference for both teachers and learners. Thus, this research sought to introduce a diagnostic process within a classroom context, with a tool that could be embedded into the teaching process and which had the potential to take the field of diagnostic assessment forward, at least in the classroom. It was with these thoughts in mind, that I decided to undertake this research project. In order to take the field of diagnostic assessment in the classroom forward and in effect this research project,

a language skill needed to act as the vehicle which required diagnosis. Thus, this study set out to investigate diagnostic listening assessment and to explore how feasible such a process would be in an English language academic setting in Saudi Arabia. According to Harding, Alderson and Brunfaut (2015), research into listening assessment has a "Cinderella" (p.326) status amongst the four skills of reading, writing, listening and speaking. They state that whilst the process of diagnostic assessment they have produced has a strong theoretical basis, "its application in the field of language assessment remains untested" (p.318).

This study sets out to explore how feasible it is to implement some of the diagnostic process and principles as set out by Harding et al. (2015) as shown in the diagram below (p.319).

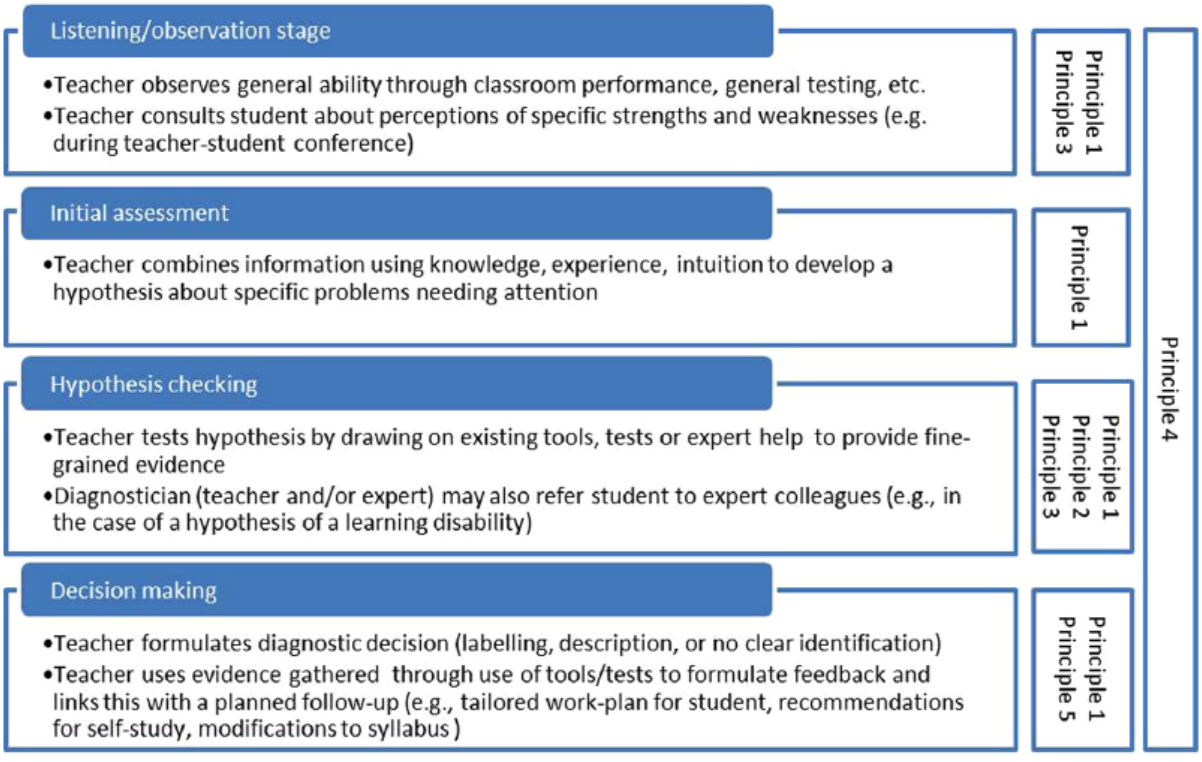


Figure 1 Diagnostic Process (Harding, Alderson, & Brunfaut, 2015)

The main objective of this research is to draw on some aspects of the model above, and ascertain whether these particular stages or principles are applicable as they currently

stand, or whether the model needs to be built upon or modified to take into account the real teaching context, based on evidence from this research project.

It seeks to explore and examine how practical, feasible and implementable diagnostic listening assessment is, especially to see if it is feasible to collect diagnostic data as part of the normal classroom lesson and also to try to gradually build a learner and class profile with the aim of creating a greater awareness especially for the teacher in relation to individual differences (strengths and weaknesses) that may exist amongst the learners. The research also seeks to address a problem raised by Field (2008), that the existing listening lesson format, largely follows a testing approach, where a listening passage is played, responses answered, and answers checked without the teacher being able to ascertain for sure why some learners may not have been able to answer correctly the tasks at hand, indeed even why they got the correct answer. At the start of the research, it was hoped that the findings would also help to create a bridge between diagnostic theory and teacher practice in the classroom.

1.3 The significance of this study

Diagnostic assessment when compared to mainstream language assessment has been relatively less researched than other areas. In their article, Harding et al. (2015) raise important points that need to be researched in order to further develop and expand this area. This study therefore I believe is important from the point of view that it tries to build upon and add another perspective to this important area.

1.4 Introducing the research focus

This research investigated whether diagnostic listening assessment was possible on a regular basis in the classroom, the extent to which specific learner problems could be

measured when learners were involved in the process of listening and trying to respond to listening tasks, and the usefulness of the overall class and individual profiles produced as a result of collating learner responses from a Listening Review Sheet (LRS) that was used during the listening lesson, and which had been created by the researcher.

1.5 Overview of the Research Design

Diagnostic listening assessment is an under-researched area, and is gradually being addressed, though mostly theoretically (Harding et al., 2015). As a result of the limited research in the area (Harding et al., 2015), the current research study followed an exploratory approach.

The research was conducted over a six-week period. The approach attempted to take into account Harding et al.'s, (2015), diagnostic process (see figure 1 above). The following table highlights the extent to which implementation of this process occurred.

Table 1

Aspects of diagnostic framework implemented during research

Listening/observation stage
Teacher observes general ability through classroom performance, general testing etc . • General observation and use of tests : Listening Review Sheet (LRS) (3.8.3 & 3.10.1) *Teacher consults student about perceptions of specific strengths and weaknesses (e.g. during teacher student conference)* • Student perceptions regarding perceived strengths and weaknesses: Interviews/teacher-student conferences (3.8.5 & 3.10.2).
Initial assessment
Teacher combines information using knowledge, experience, intuition to develop a hypothesis about specific problems needing attention

- Based on data from the LRS (3.8.3 & 3.10.1), class profiles (3.10.1.1 & 4.2) student profiles (3.10.1.2 & 4.3) and interviews (3.8.5 & 3.10.2), specific learner problems were identified.

Hypothesis checking

Teacher tests hypothesis by drawing on existing tools, tests or expert help to provide fine-grained evidence.

- Not tested, but multiple data collection cycles through the use of the LRS resulted in profiles which helped to corroborate the kinds of problems learners seemed to have (3.10.1.1 & 3.10.1.2). However, these were at a basic level and not fine-grained.

Diagnostician (teacher and/or expert) may also refer student to expert colleagues (e.g., in the case of a hypothesis of a learning disability)

- Outside scope of research

Decision Making

Teacher formulates diagnostic decision (labelling, description, or no clear identification)

- Formulating diagnostic decisions (labelling, description, no clear identification) which were already built into the LRS, resulted in learner profiles (4.2 & 4.3).

Teacher uses evidence through use of tools/tests to formulate feedback and links this with a planned follow-up (e.g. tailored work-plan for student, recommendations for self-study, modifications to syllabus

- Outside scope of research

In addition to the diagnostic process, was the need to take into account the (i) knowledge, and (ii) experience, for any diagnostician or diagnostic process (Alderson, Brunfaut, & Harding, 2014). Knowledge is acquired through "formal study, pre-service and in-service" (p.11). Whereas tools may be available, it is experience that helps decide on when and how to use these tools. For many, 'having seen it before' (p.11), is fundamental to diagnosis. As diagnostic listening assessment is a little studied phenomenon in the classroom, this research design assumes that both knowledge building and experience are areas that need to be developed. As this research is exploratory, its purpose will be to learn, create and try out

a tool and gain some initial experience of diagnosing listening in a classroom context. The research involved some fine-tuning, reviewing and amendment as was necessary.

1.6 Organisation of the dissertation

Following this chapter, is chapter 2, which is a literature review of diagnostic assessment, as well as a review of how the issue of sub-skills in listening have been described, and how these can possibly be adapted to suit the current project. Following on from this, chapter 3 describes the research methodology and design, describing the processes that took place in conducting the research and collecting the data. In chapter 4, the results produced are discussed, which include class and individual profiles, reporting students responses in relation to their LRS responses, as well as descriptive statistics and a lexical analysis of the listening passages. This chapter includes discussions relating to the results. Chapter 5 summarises the findings and considers future directions for research.

1.7 Scope of data and study

The table below provides a table summarising the key sets of data produced during the research project:

Table 2

Key sets of data produced during the research project

Tool	Data produced	Product of data
Listening Review Sheet (LRS) used in classroom during listening lessons to provide diagnostic information.	Learner perceptions relating to: **Listening passage:** Understand general idea Topic interest Percentage understood Speed perception New vocabulary	Class Profiles. Learner Profiles.

	Vocabulary known but forgotten **Main idea and detail section tasks** Learner perceptions of difficulty in relation to listening tasks	
Learner Interviews	Further insight into responses on LRS	Additional data for class and learner profiles.
Descriptive and inferential statistics based on LRS responses.	Data relating to learner responses on LRS	Correlation of learner perception data with other LRS variables.
Lexical analysis of listening passages	Lexical measures	Comparison and correlations with LRS data.

The main and pivotal source of data was the Listening Review Sheet (LRS), which acted as the conduit for collecting diagnostic data. The data from the LRS, which was used during every listening lesson, resulted in diagnostic data at both class level, as well as individual level, allowing the teacher to be able to then ascertain where learners seemed to be having difficulties. Interview data provided further insight into learner responses. The result was profiles which provided information about learner strengths and weaknesses at an individual level, as well as problems at a class level with particular listening passages.

Descriptive and inferential statistics relating to the responses on the LRS provided data about the relationships between learner responses and the passages concerned. Finally, listening passage lexical characteristics were measured, and compared to learner responses on the LRS to ascertain if there were any consistencies between learner perceptions of the listening passages and these lexical measures.

The aim of the data was to empower the teacher to ascertain learner strengths and weaknesses, and to become aware of the listening problems learners encountered, as well as problems faced when dealing with the accompanying task. The research did not endeavour to investigate possible remedial action; rather, it focussed on the task of obtaining diagnostic data.

1.8 Structure and Timeline

The table below provides the timetable for the research:

Table 3

Research Timetable

Process of research	**January 2016 – March 2016**
Recruit learners for research	
Implement research tool in classroom	
Interview participants	
Writing up phase	
Review and analyse of data	
Literature review and writing up of dissertation	**April 2016 – May 2017**

Chapter 2 LITERATURE REVIEW

2.1 Introduction

Diagnostic listening assessment is a sub-area of research within the diagnostic assessment field and possibly the least researched aspect of it, at least in the classroom. This seems to run parallel to listening research generally, which is also under researched compared to other skills (Alderson et al., 2014). In this chapter the discussion revolves around the concept of diagnostic assessment and reviews the major literature related to it. I then take a more specific look at proposals about what diagnostic listening assessment could be (Harding et al., 2015), followed by a look at how attempts have been made to break-down the listening process into micro-skills, and a discussion of Field's (2008) ideas on the process of listening.

2.1.1 What is diagnostic assessment?

In the language assessment literature, diagnostic assessment has been defined in different ways.

All tests are potentially useful for diagnosing strengths and weaknesses, but more often than not, this refers to a test that is specifically designed for gathering specific information about a specific domain (Bachman, 1990).

Often, these tests are created to discover areas in which a learner might need help. Even proficiency or achievement tests may be seen as having some usefulness in trying to diagnose at a general level, or other tests which focus more on the finite details of a language skill, and thus are more specific. However, they are seen as difficult to create, and so more often than not, other tests, meaning those that are not specifically created for diagnosis purposes, may be used for the purpose of diagnosis (Alderson, Clapham, & Wall, 1995).

Diagnostic assessment is also seen as part of formative assessment, for example where teachers may assess learner's strengths and weaknesses, or through other means, such as through the use of commercially produced or 'special' tests (McKay, 2006). Diagnosis is also seen as something that occurs at the beginning of a course, known as 'initial diagnosis', in order to ascertain strengths and weaknesses. This is then followed by 'ongoing diagnosis', at particular points in the learning continuum. At this point, appropriate feedback can be given, and teachers can decide what needs to be done next (McKay, 2006). With regards to what 'initial' and 'ongoing' tests would look like in terms of specificity, or the type or level of information, there is no explanation given by McKay. When discussing grammar tests, she mentions diagnostic assessment in the sense of discrete items. Tests that are primarily diagnostic in nature and low stakes, need not have as high a reliability as those that are more high-stakes in nature (McKay, 2006).[1]

Diagnostic tests have also been said to be for the purposes of seeing how students are doing on a particular language programme, probably created by teachers, with the intention of seeing to what extent programme objectives have been achieved (Brown, 2005). Thus, the primary purpose is to see how learners perform in relation to curriculum objectives, see where they are performing well or not so well, and then to focus on the areas that they are having problems with (Brown, 2005). It is suggested that diagnostic tests would be conducted at the beginning and middle of a course of study and thus would be criterion-referenced, and would be used to decide whether student placements were appropriate, in addition to looking at student strengths and weaknesses (Brown, 2005). Thus diagnostic tests would be used at intermittent periods. Criterion-referenced tests, which would test particular objectives, and

[1] Test reliability refers to tests where the test scores are likely to be consistent. For example, if a teacher administers a test to students on a particular day, and then a few days later the test is administered again, the scores would be expected to be close or similar to the previous weeks scores. This contrasts with unreliable tests, where scores would not be consistent and thus vary widely (Bachman, 1990; Brown, 2005).

thus have a diagnostic quality about them, could also be used at the end of the period and at the same time, take on the role of an achievement test (Bachman & Palmer, 1996; Brown, 2005). No further detail is given concerning what such diagnostic tests would look like or what kind of qualities they would have except that they would be provided by the administration, and that there may be situations where lead teachers and other colleagues may need to modify the tests with colleagues. Interestingly, recent research indicates that teachers seem to be ill-equipped in terms of the skills and knowledge required, to be able to produce even tests in general (Fulcher, 2012; Hasselgreen, Carlsen, & Helness, 2004; Voght & Tsagari, 2014). Therefore, expectations from teachers to create or modify relatively complex diagnostic-type tests may be somewhat ambitious

Although diagnostic tests are designed to assess specific elements of language, Brown (2004) suggests that in some contexts, diagnostic tests and achievement tests are 'indistinguishable'.

Some Language Testing and Assessment books do not seem to make reference to Diagnostic Testing such as McNamara (2000) and Fulcher and Davidson, (2007), which confirms Alderson's (2005) view, that the literature on language assessment either ignores or provides very limited guidance on how diagnosis should be conducted, what the contents of a diagnostic test should include, or any kind of framework, theoretical or otherwise upon which some kind of beginning can be made.

Based on what has been discussed thus far, diagnosis or diagnostic testing seems to encompass the following qualities:

1. Diagnostic tests are used to determine strengths and weaknesses (Bachman, 1990), and thus used to determines areas for which learners need help (Alderson et al., 1995).
2. Any test (for example proficiency, achievement and placement) can potentially be used for diagnostic purposes too (Bachman, 1990; Alderson et al., 1995; Bachman &

Palmer, 1996; H.D. Brown, 2004; J.D. Brown, 2005). Thus, despite the main purpose of a test being clear, a test is still seen as having potentially a secondary purpose, i.e. diagnosis.

3. Test items are discrete (McKay, 2006) and also difficult to create (Alderson et al., 1995).
4. Diagnosis is low-stakes and thus reliability not as important as on high-stakes tests (McKay, 2006)
5. In terms of timing, a diagnostic test can be conducted at the beginning, during, or at the end of a period of study as well as on an ongoing basis (Brown, 2005; McKay, 2006).
6. Ideally includes feedback (McKay, 2006).

Thus there are varying views of what diagnostic testing really is.

The first detailed attempt at describing diagnostic assessment in terms of foreign language learning was by John Alderson in his seminal work 'Diagnosing Foreign Language Proficiency' (2005).

In discussing the idea of what diagnostic assessment could be, Alderson talks about the underlying and fundamental need to 'help learners make progress' (Alderson, 2005, p.1), however, he expresses concern that there is a lack of knowledge about how language develops amongst the various stakeholders such as testers, researchers, and theoreticians, and thus the chances of being able to help learners progress is at the current time limited. In his view, both high-stakes testing specialists and classroom assessment specialists have neglected to investigate or research learner strengths and weaknesses. He does suggest this may be due to the greater interest in and influence of high-stakes international exams on the research agenda. Although his book is based around the Dialang test (Alderson, 2005), an online

computer-based test in multiple languages that tests all four language skills, his purpose is to generate a debate on the issue of diagnostic testing.

After analysing various definitions and synthesizing the testing and assessment literature, he also comes up with a list of qualities based on his own synthesis which he suggests could be used to guide thinking on the way forward. In addition to the points already listed above about diagnosis and diagnostic testing, he lists items (Alderson, 2005, p.11-12) such as: detailed feedback, low anxiety due to the low-stakes nature of diagnostic testing, that tests should be based on language theory rather than a global theory, they are likely to be less authentic, that the focus is not language skills, but rather language and that the focus is likely to be on low-level skills, and he gives the example of phoneme or grammar tests as being more useful for diagnosis rather than for example vocabulary (Alderson, 2005).

Diagnosis it seems means different things to different people, depending on the kind of context and purpose within which any so-called 'diagnosis' takes place. If only language is to be diagnosed, should language skills be excluded, when they play a part in the development of language. If the focus as Alderson states above should be on only lower level skills such as phonemes and grammar, how useful is this when learners are still able to perform well on language proficiency tests. One could ask why we should exclude vocabulary from diagnosis. Depending on the context, ascertaining learner strengths and weaknesses in vocabulary knowledge could potentially help language learners progress through targeted and specific remedial work (Read, 2008; Urmston, Raquel, & Tsang, 2013). With regards to how fine-grained diagnosis should be, this would also depend on the context and purpose. Certainly, within a classroom (the focus of this research), this would not be a straight forward undertaking. If we see language learning as a complex intricate whole, then perhaps we need to be open to a view of diagnosis at different levels. If we take an onion as an example, it has many layers. As each layer is peeled away, another layer is revealed. This

process continues until the core is reached. However, each layer is probably essential for the whole. In this sense, focussing on how finite diagnosis should be, may not take into account the overall complexity of language development. In terms of validity, perhaps new definitions are required that take into account context and purpose, and the complexity of language learning.

In essence, there seems to be much variation in what the literature says about diagnosis and diagnostic testing. This variation could perhaps be lent some coherence by defining diagnosis in the following way:

Diagnosis in language learning is first and foremost the discovery of strengths and weaknesses, and then the alleviation of learner weaknesses through any means available. What is to be diagnosed and how infinitesimal it is, depends upon the diagnostic context and purpose.

Alderson (2005) believes there has been a failure to create diagnostic procedures that are applicable generally, or which are applicable to "one-to-one individualised procedures for diagnosis" (p.25). In effect, these issues need to be addressed. He summarises by saying that, "Only through the trial and error of developing diagnostic instruments, based on both theory and experience of foreign language learning, are we likely to make progress on understanding how to diagnose, and what to diagnose" (Alderson, 2005, p.25).

An attempt to define diagnostic assessment or as he refers to it as diagnostic language assessment (DLA), is made by Lee (2015) who defines it as: "to be the processes of identifying test-takers' (or learners') weaknesses, as well as their strengths, in a targeted domain of linguistic and communicative competence and providing specific diagnostic feedback and (guidance for) remedial learning" (p.303). Thus Lee believes that the major components of diagnosis are 1) identifying strengths and weaknesses, 2) feedback, and 3) remedial learning.

Henceforth, we will refer to Lee's term Diagnostic Language Assessment or DLA to encapsulate the idea of diagnostic assessment. Lee's definition of DLA breaks down diagnostic assessment into three broad components that have the potential to act as overarching labels that describe the process of diagnosis. Harding et al's (2015) diagnostic process also falls easily into these three components. In terms of the current research, these components also help to clarify where this research starts and ends, which is within the first component – identifying strengths and weaknesses.

Before moving on to a closer look at the way diagnostic information is produced, we look at the different contexts within which attempts have been made to realise DLA.

2.2 Contexts of Diagnostic Language Assessment (DLA)

There have been a variety of differing contexts within which attempts have been made to implement DLA. Looking at the different contexts will enable us to understand the continuum that currently exists in the field of diagnostic assessment.

2.2.1 Large scale diagnostic tests

Dialang (Alderson, 2005) is probably the most well-known diagnostic testing system, possibly as a result of Alderson's (2005) publication relating to Diagnostic Assessment. A computerized system, it is open to anyone who has access to the internet, and is a collaborative European wide project. A fully computerised diagnostic system, it is based on the CEFR and assesses reading, writing, listening, grammar and vocabulary in 14 European Languages (Alderson, 2005; Alderson, 2007). The tasks were created using detailed Council of Europe task and test specifications. In terms of reporting, it provides a CEFR (Alderson, 2005) level, thus indicating a range within which someone's language ability falls. In addition to this, a profile report is created immediately afterwards, which details how well

someone has done for sub-skills in each of the macro skills. The profile can then be used by the learner to examine areas of strength and weaknesses based on this profile. Dialang also encourages self-assessment in the form of Can-Do statements related to the CEFR. Participants may then compare their responses to the actual results created by Dialang, which is meant to encourage reflection on why such differences may occur (Alderson, 2005). In his examination especially of the results of the English pilot version of the test, Alderson notes that for example for reading, although learners may be categorised at a higher CEFR level, they all still seem to be weak in particular sub-skills. Thus, as he notes, it seems that inferencing as a sub-skill may not necessarily improve as one's reading ability develops, like for example understanding the main idea. He also noted in the listening section, that low level learners were also able to respond to some inferencing questions as well as main idea type questions. So there was it seems some limit as to how useful the diagnosis was. It is suggested that this is indicative more of a problem or limitation of the CEFR. It's focus on communicative language or on the idea of communicators as 'social agents' (Alderson, 2007, p.26), focuses on certain types of language and output, and overlooks other aspects of language. Alderson goes on to suggest, that rather than focussing on what learners 'Can-Do', may be there needs to be a look at what learners cannot do yet (Alderson, 2007).

In New Zealand, the Diagnostic English Language Needs Assessment (DELNA) system was one of the first diagnostic systems to be introduced within a university context in 2002 (Doe, 2014; Read, 2015), which was and is used to help with the diagnosis of student language problems due to an increasingly diverse linguistic student population (Read, 2008; Dunworth, 2009; Murray, 2014) . DELNA was divided into Screening (computerised) and Diagnosis (paper-based), with Screening (Elder & Randow, 2008) assessing vocabulary and reading speed, in order to, according to Read (2008), quickly ascertain native from non-native speakers of English. Those who attained a minimum score on the first section would then be

exempt from the next part known as the 'diagnosis' (Read, 2008), which comprised listening to a mini-lecture, reading academic-type texts, and writing (Knoch, 2009) an interpretation of a graph. The diagnosis phase takes two hours and resembles an international proficiency test, with a 30 minute listening test requiring responses to multiple choice and information transfer to an answer sheet, one or two reading texts with a variety of response type items, and writing tasks requiring 200 words. This was offered across the university. It was not used for admissions purposes, was low-stakes in nature, and gave students the option to use the results to get more academic language support, which was the primary purpose of the system. In a way, it was used to place students into courses to enhance their academic literacy by identifying the strengths and weaknesses and then providing them with support. In 2007 over 5,000 first year graduate students took the test. The results were reported in a diagnostic report at a sub-skill level. This type of assessment is also known as Post-entry (or post-enrolment) language assessment (PELA) (Doe, 2014) as it's focus is on those who are already at university. Overall, the post-entry test has been successfully implemented (Read, 2015).

Following on from the example in New Zealand, other systems have also been introduced. One such system is the online diagnostic assessment system (OAES) at a university in Taiwan (Yin, Sims, & Cothran, 2012) which has a general proficiency section which assesses listening, reading and grammar with 60 questions per skill and which provides 'macro-diagnosis' (Yin et al., 2012, p.3). In addition to this general proficiency test, there are three other tests for each of the same skills, but providing 'macro-diagnosis'(Yin et al., 2012, p.3) based on specific constructs. It is fundamentally a low-stakes pedagogic tool with apparently low uptake (Yin et al., 2012) as a result of the voluntary nature of taking the test and or the perceived lack of urgency of the test. Many of the sub-constructs for the grammar test are drawn from the Dialang System (Yin et al., 2012).

In Japan, the EDiT or English Diagnostic Grammar Test (Koizumi et al., 2011) focuses on particular aspects of grammar and is used to test secondary school students. Primarily it focuses on areas that a teacher may not be able to detect in class, and results in a report that could be used by teachers to help them with their teaching planning.

In Hong Kong, The Diagnostic English Language Tracking Assessment (DELTA) was developed to diagnose listening, reading, grammar and vocabulary (Lockwood, 2013; Urmston, Raquel, & Tsang, 2013) at university level. The web-based system assesses the four skills using MCQ test items. In order to be able to provide detailed information on test performance, the development of the test included a literature review of taxonomies resulting in lists of sub-skills. These were then reduced in order to make the test writing manageable and to ensure that the test report was easily comprehensible to users and stakeholders such as students and teachers. The reading and listening subskills were finalised based on consultation between test item writers, and experienced teachers. It was ensured that the subskills linked directly to how the learning materials were classified. The grammar test items were based on actual learner errors, making these useful for further targeted and remedial study, with a focus on diagnosing grammatical problems when writing essays, rather than isolated grammar structures. The vocabulary test was based on the Academic Word List (AWL) by Coxhead (2000), in order to reflect the target vocabulary used at tertiary college level. The availability of materials related to the AWL for further learning also aided in the decision to use this as part of the DELTA diagnostic system.

As can be seen, these are highly intricate, well-developed systems that focus on diagnosing problems across a large number of students from a variety of contexts. A question that could arise is whether these systems properly cater for more individualised problems.

2.2.2 High-stakes proficiency exams

In terms of research, attempts have been made to retrofit, create and extract diagnostic information from tests whose primary aim was language proficiency. Thus, several attempts have been made to extract diagnostic information from the TOEFL ibt (Sawaki, Kim and Gentile, 2009; Jang, 2009a; Jang, 2009b; Lee & Sawaki, 2009; Kim, 2011; Sawaki, Quinlan, & Lee, 2013), the Trinity GESE (Révész & Brunfaut, 2013), MELAB (Li, Hunter, & Lei, 2016) and the PTE (Huhta, 2014; Brunfaut & Révész, 2015).

Cognitive Diagnostic Assessment (CDA) models have been used extensively in an attempt to produce fine-grained information about learners based on test items, and thus create learner skill profiles (Jang, 2009). An example of the use of a CDA model is Jang's (Jang, 2009) attempt to produce diagnostic information from LanguEdge (Jang, 2009), an instructional software developed for teaching and instruction soon after the release of the TOEFL (iBT).

Although LanguEdge was not a diagnostic system, the research conducted by Jang applied the Fusion statistical model to the reading component of the software. The Fusion (CDA) statistical model was used to break-down further the reading test item responses into micro-skills. The purpose was to measure "learner competence on an array of cognitive skills" (Jang, 2009,p.36), so that strengths and weaknesses could be reported at a more detailed level. In effect, this reflected a test item's diagnostic ability.

The relationship between cognitive skills and test item values is expressed in a table known as the Q-Matrix. Thus, an item is given a value based on the kinds of cognitive skills that may be required to respond to an item correctly. In this sense, the Fusion model tries to predict responses based on these parameters (Jang, 2009). A main conclusion from the research was that retrofitting a CDA model to a system where the tasks are not specifically

created for diagnostic purposes restricts and hampers the kind of information that can be created.

There have also been other studies into diagnostic assessment that have incorporated the Q-Matrix (Buck & Tatsuoka, 1998; Sawaki, Kim, & Gentile 2009; Jang, 2009; Lee & Sawaki, 2009).

The use of the Q-Matrix and other CDA models have been criticised by Alderson (2010) and Davidson (2010), primarily because diagnostic assessment is ideally seen as something that should be accessible to teachers in the classroom who may not be familiar with such methods. Furthermore, some of the research incorporating such models has been retrofitted to tests whose purpose was never diagnostic in purpose. However, it should be said that as diagnostic assessment is seen as something that is in its infancy, such efforts still add value as they extend the idea of what diagnosis is, and also contribute to the discussion about what is and what is not diagnosis.

2.2.3 Classroom related research

There have been a limited number of diagnostic research studies which attempt to relate directly to the classroom context (Fox & Hartwick, 2011; Hirai & Koizumi, 2013; Jang, Dunlop, Park, & van der Boom, 2015) using a diverse set of assessment tools and models.

Hirai and Koizumi (2013) developed and compared three diagnostic speaking rating scales in relation to a task called the Story Retelling Speaking Test (SRST) for use in the classroom. The three rating scales were: 1) A four-criteria, binary-choice, boundary-definition rating scale that had been empirically derived, 2) a three criteria scale modified from the first, and 3) an analytical multi-trait rating scale that followed a conventional analytical format (Hirai & Koizumi, 2013, p.398). The multi-trait scale was deemed more

practical by the teachers, despite the second rating scale having been shown to be more statistically reliable and valid. Although one of the conclusions reached was that format impacts on the qualities of a test, an important finding was that practicality is important in order to be able to make any diagnostic assessment in the classroom a feasible exercise.

Jang et al. (2015), investigated the impact of diagnostic feedback on young classroom based learners using a "Holistic Diagnostic Feedback" (HDF) (p.359) report, to ascertain how learners identify with, and sift through the feedback based on their own inclinations and aims, and how these may assist or impede the potential advantage of feedback in their take-up of judgements as well as potentially shape their future trajectories towards further learning. Based on a localised reading test, the report included three types of information. Firstly, a list of reading subskills derived from a CDA analysis; secondly, responses to a learner self-assessment questionnaire about how learners perceive their own abilities; and thirdly, a "goal orientation profile" (Jang et al., 2015, p.359), the data for which was based on responses to questionnaires by both learner and parent. Jang et al. define goal orientation as a "set of goals"(p.361), which it is believed makes diagnostic feedback more useful as feedback can indicate the extent to which progress has been made towards the goals. It was found that learners generally rated themselves more highly than their actual performance which was consistent with previous research (Jang et al., 2015). Taking these considerations into account, Jang et al. concluded that learner feedback needs to be understood through the eyes of learners in order for this phase of diagnostic assessment to be successful.

Fox and Hartwick (2011), experimented with a diagnostic portfolio within a university EAP program which was run twice a week over 12 weeks. Learners represented a variety of nationalities and linguistic backgrounds who were planning to complete undergraduate degrees in different majors. They were given a diagnostic test based on the Canadian Academic English Language (CAEL) Assessment near the beginning of the course,

and then divided into groups based on learner profiles (Fox & Hartwick, 2011) that reflected their test results. Initial scores for reading, listening and writing were then broken down into subskills based on individual test items, including the writing component which was marked by CAEL trained raters. Additionally, responses to a self-assessment questionnaire were also compared to actual results, indicating where learner self-assessment had been similar or different to their actual test results. The purpose of the latter was to raise learner awareness. Learners were put into small groups and helped by a teacher for an hour and a half a week on reflections and activities that focussed on their specific issues, with revisions and changes as appropriate in order to drill-down (Fox & Hartwick, 2011, p.50) as much as possible to the problem at hand. A group of 5 MA student volunteers created and analysed tasks and specifically created diagnostic tests and related activities, monitored language development and met to discuss problems and issues under the direction of the researchers. In effect, there were two researchers, two teachers and five research volunteers working on the project, trying to implement the diagnostic portfolio.

On a weekly basis, learner problems were identified and treated using specially created tasks with skills being continually diagnosed. Initial results seemed to indicate that as learners became more aware of the subskills they found it useful to know where they had problems. Both learners and teachers found the diagnostic information revealing and interesting as these sometimes contradicted self-perceptions of what were perceived as strengths and weaknesses. The information relating to listening subskill weaknesses was particularly revealing to the teachers. This was to be expected as the skills are receptive and not easy to assess in normal circumstances. With quick feedback on in-house diagnostic tests created by the volunteers, and diagnostic tests being geared to individual learners, the whole portfolio/diagnostic process began to have meaning and purpose for the learners. The

diagnostic tests, tasks and activities then became the content of the portfolio, demonstrating a record of achievement.

Whereas there were limitations in ascertaining all of the problems at subskill level as a result of the non-completion of the initial diagnostic test due to length, two important limitations were noted. Firstly, a relatively large number of people were required to make the process feasible. Secondly, the amount of time and effort taken to support, discuss and create activities, tasks and tests was significant.

These last two points are important points to bear in mind and they confirm what has been said in the literature (Alderson, 2005) about the perceived time and complexity involved in the diagnosis of language problems.

However, there are nonetheless examples where diagnostic assessment has taken place at a more simple level in the classroom in which a teacher was responsible for the diagnostic process from start to finish. An example from Brown (2004), illustrates what a diagnostic test looks like by referring to a test of oral production attributed to Clifford Prator (1972, cited in Brown, 2004, p.47) in which pronunciation was assessed, by asking students to read a passage multiple times. After having read the passage multiple times, the teacher would tick boxes on an inventory list that was based on various sub-categories of pronunciation, which were then each broken down further, the results of which were then used for further action by the teacher. No further detail is provided.

Looking at the examples of attempts at diagnosis presented so far, there are characteristics that they all seem to have in common. In terms of developing tools that may help in extracting information that relate to strengths and weaknesses, no teachers were directly involved it seems in developing them. This reveals potentially a major reason for the lack of development of diagnostic assessment as a field more widely, and which is reflected in the literature generally. Instead, it is non-teaching researchers who seem to propose and

create the diagnostic tools and teachers who are then used as part of the process of 'testing out' the tools. In Hirai and Koizumi (2013) and Jang et al. (2015) diagnostic tools were developed which later on were used by teachers. In Fox and Hartwick (2011), the diagnostic tools and activities were primarily created by volunteers under the guidance of the researchers, requiring more people in addition to the teacher. Examples of where a teacher is conducting or implementing any aspect of diagnostic assessment on their own is rare, the only example provided in the literature being Clifford Prator (1972, cited in Brown, 2004, p.47), mentioned earlier.

2.3 Summarising the diagnostic assessment literature

Being able to produce fine-grained information (Sawaki, Kim, & Gentile, 2009) or rather information that can help to identify strengths and weaknesses of learners is the first step towards making diagnosis achievable. Whereas terms such as micro or macro (Alderson, 2005; Jang, 2009; Lockwood, 2013) have been used, there is still no clarity or agreement on how fine-grained or minute such data needs to be. Lee (2015) has rightly suggested that specificity and granularity are abstract concepts and the questions of how specific is specific enough is far from straightforward.

As has been demonstrated above, there are multiple streams of research related to diagnostic assessment, however, Davidson (2010) expresses concern at the gulf that is created between the complexity of some ideas (he specifically talks about CDA). For him, in addition to the focus on high-stakes testing, this results in a kind of dichotomy, oppositeness, and complexity, resulting in distance from the classroom. There is, in his opinion, a need to close or bridge the apparent gap between high-stakes testing and the classroom. The question Davidson (2010) has, is how that gap can be bridged. Alderson (2007) also reiterates that further exploratory research is needed in terms of formative assessment in both language and

general education in order to further develop what is meant by diagnostic assessment (Alderson, 2007).

Having looked at the contexts of diagnostic assessment, we now turn our attention to the listening literature.

2.4 Breaking down Listening

We now discuss attempts made in the listening literature to describe the components and elements of listening which may have the potential to be used for diagnostic purposes in terms of this research project.

2.4.1 Sub-skills Approach

The first detailed attempt at trying to describe the components or sub-skills of listening was by Richards (1983). Richards presented a taxonomy of listening micro sub-skills for conversational listening (Appendix C) and academic listening (Appendix D). This was a suggestive list which Richards presented as having potential for curriculum development. The conversational listening sub-skills were divided into thirty-three sub-skills, whilst the academic listening sub-skills numbered eighteen. He based his list on aspects of spoken discourse and after analysing listening processes.

In addition to using these micro-skills for curriculum development, Richards suggests that they could also be used for diagnostic testing. He suggests that using these in combination with rating scales, for example developed by Brindley (1982, cited in Richards, 1983), would result in a detailed learner profile. Based on Brindley's rating scale descriptors (See Appendix AM), Richards believes that if used with a "skills taxonomy" (Richards, 1983, p.231), problems with specific micro-skills could be identified, thus requiring remedial work.

Harding et al. (2015), have reservations about Richard's micro/sub-skills approach as a way to diagnose problems in listening. They believe that he does not demonstrate how these would be practicable, and that the only "unifying principle" (p.328) are contexts of listening, which provide broad labels that do not assist in terms of understanding the underlying problems, that are more atomistic in nature; a view shared also by Field (2008). For example, as Harding et al. (2015) suggest, "of what practical use is it for a teacher to know that a student needs more help with 'listening for general ideas" (p.327). In contrast to Richard's a priori list of sub-skills (Field, 2008), Harding et al. (2015) believe that process models of listening comprehension are probably more suited to designing appropriate diagnostic tools.

2.4.2 A Process Approach

In contrast to Richards, Field (2008), believes that a more concrete method of determining what constitutes listening abilities is required for example by targeting particular listening behaviour that has been observed and researched. Thus, for Field, this becomes the basis for differentiating between a sub-skills approach and what he calls the process approach. Whereas the sub-skills approach is based on assumed underlying listening abilities, he suggests an approach to breaking-down listening based on existing research and observations of listener behaviours. This is based on L1 listening research. He thus introduces a taxonomy which he divides into "decoding processes", and 'meaning building' processes (Field, 2008). Thus, listening involves recognising sounds and words – the decoding processes (Appendix E), also known as bottom-up processing, and the meaning building processes (Appendix F) – that is processes that help to contextualise and understand the message, also known as top-down processing.

Field indicates that these behaviours (decoding and meaning-building) could form the basis of a listening curriculum, allowing teachers to focus on these areas. Conversely, the

separation of listening into decoding and meaning-building could potentially be a basis to break the listening process down further, and potentially become the basis for diagnosis too. However, to date, there are still no concrete examples of how this can be implemented.

As part of the development of the TOEFL iBt, a listening process was also proposed (Bejar, Douglas, Jamieson, Nissan, & Turner, 2000) that would underlie the proposed academic listening exam. According to Bejar et al. (2000), listening comprehension consists of the listening stage and the response stage.

During the listening stage, there are three processes that are activated in real time as the sound signal leaves the speaker and is received by the listener. These are:

- Situational Knowledge (SK): this relates to context.
- Linguistic Knowledge (LK): which refers to "grammar (phonology, vocabulary, morphology and syntax)"(Bejar et al., 2000,p.2), and discourse, pragmatics.
- Background Knowledge (BK) relates to cognitive demands. This includes background knowledge, inferencing, ability and memory (p.2)

Differences in these areas will result in a different set of propositions (PR), in other words, a representation of the listening in the mind of the listener. This mental representation or set of propositions are then accessed by the listener to respond to a set of questions. If there is predominantly new information in the aural stimulus, then more time will be spent trying to sort out or order the information, thus potentially impacting on the development of the listener's version of what is being heard. The type or mode of the response required, for example, written or spoken, will also impact on the quality of the response. The quality of the representation or set of propositions (PR) is based on individual knowledge and cognitive factors. All of these factors interact with each other in the process of listening and responding.

The process outlines also mirror very much what occurs in the listening lesson classroom. Rost (2011), breaks down the goals and cognitive processes that learners are said to have, and which are detailed in Jamieson et al. (2000). Rost's break-down is represented in table 4 below clearly showing what the goals and cognitive processes of the examinees are during the academic lecture process, as well as the variables that could impact each goal.

Table 4

Modelling listener processes during assessment recreated in a table format (Rost, 2011, p.219-221)

Stage 1: Listening to the Stimulus	
Goal	Listen to a talk (the stimulus) Remember information Answer questions after this
Process	Represent in working memory important info that seems important
Variables that may affect this process	**Stimulus variable:** • Length of lecture • Syntactic complexity • Density of information • Lexical difficulty **Listener Variables:** • Knowledge of the context of the task • Knowledge of the language • Attention • Working memory capacity • Background knowledge
Stage 2: Listening to or reading each questions	
Goal	Understand the questions and what is required for a response
Process	• Identify the given and requested information in the question • represent in working memory the requested information
Variables that may affect this process	**Item variables:** • Lexical difficulty • Syntactic complexity

	• Length **Listener Variables:** • Knowledge of the context of the task • Knowledge of the content of the language • Attention • Working memory capacity • Background knowledge
Stage 3: <u>Searching</u> for the correct answer	
Goal	Retrieve information from stimulus that answers the question
Process	Search working memory for information in the stimulus that matches the information requested in the questions
Variables that may affect this process	**Stimulus variables** • Length of lecture • Syntactic complexity • Density of information • Lexical difficulty **Item variables:** • Type of information • Type of match • Explicitness • Main/supporting idea redundancy **Listener variables** • Knowledge of the context of the task • Knowledge of the language • Attention • Working memory capacity • Background knowledge
Stage 4: <u>Identifying</u> the correct answer	
Goal	Select the correct answer from the options given
Process	Identify an answer to the question by finding a match with the appropriate information from working memory and verifying that none of the other options is a better match
Variables that may affect this process	**Stimulus variables** • Length of lecture • Syntactic complexity • Density of information • Lexical difficulty **Item variables**

	• Type of information • Type of match • Explicitness • Main/supporting idea redundancy • Plausibility of distractors **Listener variables** • Knowledge of the context of the task • Knowledge of the language • Attention • Working memory capacity • Background knowledge

Vandergrift and Goh (2012) also attempt to describe the listening process. They divide the listening process into Perception, parsing and utilization.

- Perception

The recognition of sounds as words or meaningful chunks of language occurs through an acoustic-phonetic processor. These remain in working memory for a short-time to process meaning. Only some sounds are retained for processing dependent on listener's language proficiency. Some sounds especially for a beginner whose L1 may not have some of the sounds of the L2 can result. Other issues that impact on the processing of the sound stream include speed, dialect, dense text on unfamiliar topics. This stage is the bottom-up phase, which becomes more efficient with greater and frequent exposure to the phonetic sounds of the L1.

- Parsing

At this stage, the sounds are broken-up or segmented based on syntactic and semantic signals that result in a mental model or representation of the combined words. The comprehension process continues as more sounds emanate from the acoustic-phonetic processor, with bottom-up processing working together with top-down processing via the conceptualiser. As meaningful chunks are created, there is a matching process with the lexicon that is in the listener's mind, with word lemmas and lexemes being used to ascertain

possible lexical options, which may be informed by context. The example of the similarly beginning words, flour and flower, are given to illustrate the kinds of decisions being made by the learner during the listening process. Perception and processing in the meanwhile are working together, until a valid mental representation is arrived at. The conceptualiser checks the result of this combined processing and matches this against what has been heard so far, and what is held in long-term memory.

- Utilisation

The result of what has been explained thus far, results in the development of a mental representation. This is known as the utilisation stage. The conceptualizer is responsible for this process, shared by both the comprehension and production processes. This is a top-down process, with resultant meaning compared to context, listener knowledge of the speaker, tone and any other relevant information such as discourse or pragmatic knowledge. The resultant meaning is the utilization. This process is active for texts of all sizes, from small, as well as large, and may also be a combination of utterances. The listener will add their own interpretation.

According to Vandergrift and Goh (2012) all of the above is controlled by the listeners, metacognition. This is the individual ability to plan, monitor, and solve or negotiate meaning.

Following on from his views on a process approach to listening, Field's ideas are developed further in the form of what he calls 'a cognitive processing framework for listening' (Field, 2013, p.93) which is introduced as a tool to validate listening exams by Cambridge Esol.

The framework is comprised of the following:

Lower-level processes

- Input decoding:
    - This relates to the sound that first emanates from the speaker and where a listener will go through the process of converting the sounds to phonemes, to syllables, word chunks, clauses and longer strings of words. Intonation and stress are also important elements in the stream of sound.
- Lexical search:
    - When words are uttered, a process of matching and selection is said to occur, where the listener is in the process of retrieving words from memory, and as more information is provided through the stream of speech, decisions are made about the probable word concerned. There are also additional processes at play that are impacting on this selection, which includes segmentation, in other words where the words begin and end, or where their word boundaries fall, which may not often be as in the written form especially during speech as words before and after may impact on the way the words are spoken, possibly changing the way the words are uttered. Simultaneously, stress, tone and prosody provide additional information, to help confirm or otherwise the listeners hypothesis about stream of speech. However, there is a deeper process that underlies the matching of lexis, which may include listener knowledge that relates to frequency of words, and the likely words they are typically grouped with, allowing for an element of readiness or priming to receive likely words. Additionally, the process of 'spreading activation' (Field, 2008, p.98) occurs, where words are stored in the mind through an 'intricate lexical network' where words are linked and where word association may occur, for example, the word football, may bring to mind words such as team, player, stadium and so forth, which could be relevant if the discourse takes a

particular direction. Thus, as the lexis is recognised for what it is in a literal sense, further information helps in building meaning.

- Parsing
    - Parsing focusses on grammar structure and word order or even disruptions (Field, 2008) in word order, all providing information to assist in clarifying what is being uttered. As the utterance continues, intonation and pitch assist the listener in ascertaining when a clause has come to an end, and at this point results in a proposition (Field, 2008) or an abstract idea, being the result of parsing, however, still subject to change depending on the route of the discourse. The completed utterance which includes a more complete intonation cycle, will also allow for an impression to be made of the speaker in terms of intentions and attitudes (Field, 2008).

Higher-level processes

- Meaning-construction
    - Up until now, a proposition has been formed in the mind of the listener, however, this is literal in its sense and requires further information to give it greater meaning, which will emanate from the listener. This further information may be in the form of pragmatic information, contextual information, semantic and inferential information.
        - Pragmatic information may relate to the perceptions of the listener about the intentions of the speaker, the speakers knowledge or other information known about the speaker that goes beyond the language.
        - Contextual information may be added based on the listener's knowledge about the subject at hand which may include general world knowledge or personal knowledge, knowledge about the situation and

knowledge about the speaker as well as a recollection of what has been said up until that point.

- Semantic information relates to what the listener may understand using his or her world knowledge by some of the words and ideas expressed by the speaker. This could be language for example that is idiomatic in nature, or where certain words are uttered, but which stand for or denote certain meanings or characteristics. For example to describe someone as a lion may suggest a particular characteristic which the listener is able to understand using their own world knowledge.
- Inferential information or inferencing will come into play where the speaker has not deemed it necessary to explicitly include certain information, relying on the listener to be able to ascertain or infer what is being referred to, which again comes about as a result of shared and assumed knowledge.

- Field also suggests that in the process of meaning-construction, there is the continuous use of anaphors, and linking these to their antecedents, in other words, pronouns may be used during a discourse, which refer to nouns that may have been encountered much earlier. Thus, this helps in linking ideas across the trajectory of the subject matter at hand; however, greater skill is required to be able to this. Although the ideas above have been presented in a particular order, this does not suggest that there is a particular sequence to these, in fact, Field suggests, that just the speakers "pragmatic intentions" (Field, 2008, p.101) may influence how the whole context is perceived, resulting in all of these processes working closely together. Thus, all of the listener knowledge referred to above is a key element to the process of

meaning construction in that they add to the decoding which occurred prior, resulting in a "mental model" (Johnson Laird, 1983, cited in Field, 2008, p.102) or "meaning representation" (Field, 2008, p.102).

- Discourse representations
    - Whereas meaning representation refers to the result of a particular utterance, discourse representation relates to the bigger picture, where multiple utterances are combined, and what has been retained in the mind of the listener. Thus, it is not everything that the listener has heard, rather it is, as Field suggests, a listener version. This listener version or discourse representation is deemed to be assembled as a result of four processes. These are: Selection, Integration, Self-monitoring and Structure building (Field, 2008).
        - Selecting information is a natural process that occurs, with the listener deciding whether information is important or not, and whether to hold on to, or to let go of ideas as the discourse continues. However, speaker and listener intentions and purposes also impact on the extent to which information is retained in terms of detail, or whether it is discarded.
        - Integration: As the discourse representation develops, new meanings are encountered, which may or may not relate to what has been said prior to these, requiring the listener to decide it's relevance and fit. Decisions may be made based on linking words or phrases that act as markers to the listener to act upon, or where listeners may need to infer connections or relevance.

- Self-monitoring: Related to integration, Field (2013) suggests that whereas new information and new meanings are encountered, listeners need to be continually checking to ensure that any new information is consistent with what has gone previously, thus, rather than making decisions immediately, they need to consider reserving judgment based on further information.
- Structure Building: As the discourse representation continues to develop, the skilled listener is able to gradually link important points and sub-points, which may be required especially where there is a complexity of information. Less skilled listeners may have a tendency to view information in a linear sense, when what is required is the ability to store a complexity of information. Field suggests that depending on the listening context, listeners may be familiar with particular discourse structures. He provides an example of lectures, where for example, a lecturer provides an introduction to a topic area and then follows a particular lecturing convention. Consistency in such a convention may aid listeners, something that Field has called for in the past (Field, 2008).

All of the above cognitive process models or depictions, have the potential to provide insight into why learners have difficulty with listening (Goh, 2000; Vandergrift, 2007). Although the framework has been laid-out in a somewhat linear manner, actual listening is complex, requiring a variety of processes to interact with each other simultaneously. In terms of all of these models as a basis for diagnosis, there is likely to be considerable overlap between the processes, especially at an initial diagnostic level. Irrespective of this, as the

processes are complex, and work in tandem, how deep one can analyse listening processes in the classroom and be able to differentiate the various elements is bound to be challenging.

We now take a look at causes of listening difficulty.

2.5 Causes of listening difficulty

In their investigation into factors that have an effect on listening comprehension (Bloomfield et al., 2010), an extensive review of the listening literature was conducted resulting in a summary of factors which were deemed pertinent to causing difficulty for L2 listeners. As this literature review was conducted in order to aid decisions relating to the revision and update of the Defense Language Proficiency Test (DLPT), the review focussed on listener, passage and testing condition characteristics, however, the affects of test item types or responses required were not covered. A summary of the main factors and sub-factors they discussed are listed below (Bloomfield et al., 2010):

Characteristics of the listener:

- Working memory capacity
- Metacognitive strategies
- Experience (with phonology, vocabulary size, prior exposure to language)
- Anxiety

Characteristics of the passage

- Length (Overall length, information density, redundancy)
- Complexity (Syntactic features, directness & concreteness, pragmatic information)
- Organisation (Orality, Coherence, Discourse markers, position of relevant information)
- Auditory features (Speaker accent, Hesitations and pauses, Noise and distortion, Speech rate)

Characteristics of testing conditions

- Time limits
- Multiple hearings
- Note-taking

In trying to ascertain ways and means by which listening may be diagnosed, the Bloom et al. (2010) summary list above provides a different perspective on how listening could potentially be diagnosed, and provides thus an alternative perspective to the process and sub-skills view of listening that was presented earlier. Some of the listening difficulties identified by Bloomfield et al. have resulted in research in which specific listening difficulties have been further investigated such as Révész and Brunfaut (2013), who looked at text characteristics, text explicitness. linguistic complexity and speech rate, and Chang, Wu and Pang (2013), who looked at text factors (e.g. lexical complexity), topic interest or relevance, and personal listening factors such as anxiety or nervousness, and factors related to speech such as speed, accent, pronunciation, as well as task characteristics.

2.6 Specificity and granularity in diagnosing listening

Taking into account Lee's (2015) view that specificity and granularity are somewhat abstract concepts especially when one is trying to define this for the purpose of what constitutes diagnosis, an attempt can be made to collate and extract points of potential diagnosis. Thus, based on what has been discussed in the literature review so far, listening has different levels of specificity and granularity and it is suggested that depending on the purpose and context, diagnosticians could decide from a variety of different methods how to put diagnosis into action. Thus, the list below demonstrates the multiplicity of methods through which listening diagnosis has occurred and could occur:

- Test item traits extracted using Q-Matrix and Cognitive Diagnostic Assessment
- CEFR – The Common European Framework
- Sub-skills approach – Richards (1983) taxonomy and the CEFR
- The Process Approach

- Characteristics of the listener
- Characteristics of the passage
- Characteristics of testing conditions
- Characteristics of tasks

With a potential array of methods to diagnose listening problems, how does one decide which method to use, especially in the classroom which is the basis of the current research. We now discuss this below.

2.6.1 Deciding on what to diagnose

In terms of what can be diagnosed, what needs to be decided is what is relevant and in what context, as well as how do we choose these (Alderson, 2007). This is likely to vary from skill to skill and possibly what the focus in text books is, as these heavily influence what goes on in the classroom. There also needs to be an acceptance that taking into account the current teaching context and teacher knowledge, there are constraints that prevent the devising or creation of innovative tools that allow for diagnosis and their implementation possible. With regards to listening in particular, by context and knowledge, is meant that teachers are following fixed conventions in the way they teach listening with little time to add a diagnostic process. Thus in typical listening classes, the first stage is often to pre-teach vocabulary and activate background knowledge through brainstorming, in order to prime learners for the listening that is about to occur. The second stage follows which is the actual listening, followed immediately after, by the third stage, which is the answering of listening comprehension questions phase. Many teachers are limited in what further they can introduce to the listening classroom as this ends the typical listening teaching cycle and they may have other time-imposed curriculum goals to meet. Furthermore, it seems that teachers need a deeper understanding of the listening process and teaching listening knowledge in

general which is still lacking (Vandergrift, 2007; Graham, Santos and Francis-Brophy, 2014; Graham, 2017). A way forward may be for teachers to "tap" (Alderson, 2007, p.29) into particular components of the skill, and then try to implement these within a diagnostic process. In fact asking learners themselves, possibly through self-assessments and interviews may help in gaining insight into what problems learners really have. As Alderson (2007) suggests, ascertaining whether a learner can cope with a certain type of text is not the goal, rather it is what creates the problems in doing so. Thus, there is scope for developing an instrument for use in the classroom, which can be fitted into the listening lesson, allowing for the collection of diagnostic information without interrupting the normal flow of a listening lesson. Additionally, and importantly, there is the possibility of creating individual learner profiles as the lessons progress, and data is collected from each listening lesson, or indeed for each listening passage.

An exploratory and iterative approach to researching diagnosis is something that is recognised as a valid way to research diagnostic assessment (Lee & Sawaki, 2009) and has been applied for example when using CDA in retrofit research exercises (Sawaki, Kim, & Gentile, 2009; Lee & Sawaki, 2009). Furthermore, in many learning contexts, a combination of attributes and skills may be required in order to complete tasks and or items (Vandergrift, 2007; Lee, 2015), thus potentially adding complexity of diagnosis.

For example, in a typical listening lesson, without knowing why the answer was incorrect, or even correct, we have no way of being able to address learner problems (Field, 2008). This is according to Field, the problem with the listening comprehension approach, which is that we assume that there is only one correct answer. May be aspects of the text were noticed by the listener that were overlooked by the materials writer / author. May be there were ambiguities in the listening question? By establishing how answers to comprehension questions are arrived at, we get a picture of learner's strengths and

weaknesses. As Field suggests, "until we have some kind of diagnostic procedures, the teacher can only continue to test comprehension, not teach it" (Field, 2008, p.82).

Based on the readings of the literature, and specifically taking into account Alderson's (2007) view about teachers tapping into particular potential points for diagnosis, as well as Lee's (2015) view that a combination of attributes and skills may be required in order to complete tasks and or items, a diagnostic listening assessment construct is proposed below.

2.6.2 Proposed construct for diagnostic listening assessment in the classroom

When a listening lesson is conducted, there is an interaction of a variety of factors. These include, L2 listener characteristics, listening passage characteristics and listening task characteristics including question prompts and responses. Thus, listening in the academic classroom encompasses a variety of elements in order to get to the output that suggests comprehension has been attained, some of which are not directly of a listening process nature. Rather, it could be argued, that comprehension or the lack thereof, was in fact as a result of listening processes, and non-listening process factors (Vandergrift, 2007). It is the combination of these that results in an output that suggests whether comprehension has occurred. Thus according to Vandergrift (2007), the teacher is only able to judge to a limited extent the factors that may have led to a correct or incorrect response.

Taking into account the aims of this research which is exploratory in nature, and bearing in mind it is an academic context, in which a standard listening text is used and in which the listening lesson structure includes listening passages and listening tasks, (in effect the non-listening factors that Vandergrift (2007) above refers to), the listening construct for the purposes of the research includes the following factors:

Listener perceptions of the listening passage:

- Interest in topic (Bloomfield et al., 2010)

- Understanding of the topic/general idea
- Percentage understood
- Speech rate (Bloomfield et al., 2010)
- Lexis/vocabulary: (Field, 2008, p.87)
    - Words not known
    - Words previously known but forgotten
- Other: e.g. accent

Listener difficulty with task prompts:

- Lack of understanding of vocabulary within task requirements (Field, 2008)
- Perceived difficulty of task

The justification for choosing these measures is primarily due to the researcher's experience of having taught listening within an academic context, which assumes that these are the kind of elements that might impact on listening in the classroom. Furthermore, Alderson (2007) suggests that teachers should "tap" (Alderson, 2007, p.29) into those factors that are deemed appropriate or relevant for the purpose of diagnosis. Thus, this proposed construct is divided into two

Firstly, it is suggested that aspects or perceptions that it is believed may impact on actual listening need to be considered. The proposed perception measures it is suggested can potentially be broadly linked to all of the processing models that have already been discussed. This is because the perception measures chosen reflect the surface level of processing. Thus, the table below shows each perception measure going through each of the levels of listening in Field's cognitive listening framework. The other models could equally be applied. In this context, Field's (2013) framework is used to illustrate what occurs.

Table 5

Potential links between listening perception measures and Field's Cognitive Validity Framework

Perception Measures	Lower level processes	Higher level processes
Interest in Topic	Input decoding Lexical search Parsing	Meaning construction Discourse representation
Understanding of the topic/general idea	Input decoding Lexical search Parsing	Meaning construction Discourse representation
Percentage understood	Input decoding Lexical search Parsing	Meaning construction Discourse representation
Speech rate	Input decoding Lexical search Parsing	
Lexis Vocabulary	Input decoding Lexical search Parsing	Meaning construction Discourse representation
Other: e.g. accent	Parsing	

Each perception measure except for the speech rate is assumed to be the result of both lower level and higher level processes. As Field already states, listening is a complex activity, where multiple processes are at work simultaneously. In this sense, the different perceptions impact on each other. As the process of listening is complex, one cannot assume that each perception measure only goes through one level. Top-down, bottom-up processing occurs along with other processes in order to get to the end-result. This is the justification for suggesting that most of the perception measures are as a result of both low level and higher level processes and are thus shown in the table above.

In addition to the listening perception measures discussed above, what also is taken into account is whether learners are in some way prevented from completing the task required, as a result of the lack of understanding of the task itself.

These areas are not easily observable by a teacher; however, any data related to these could act as valuable source of information that could potentially be actioned through remedial work. Diagnosis therefore here means, not only listener perceptions of the listening passage, but listener perceptions of the difficulty with the actual listening task. As Vandergrift (2007) suggests, listening processes are complex, and successful comprehension occurs as a result of a variety of knowledge sources, characteristics and contextual factors that need to be researched. In this sense, the construct here tries to encompass some of these contextual factors in order to aid teachers in the process of diagnosing some of these complexities.

In their major work on listening difficulties, Bloomfield et al. (2010) reviewed the literature relating to topic interest and learner ability to comprehend, speech rate and vocabulary as well as other factors. We discuss their review in the context of this research.

2.6.2.1. Topic interest and general understanding

With regards to topic interest, listening passage topics that are known to learners, are typically, easier to comprehend, topic areas that are not familiar to learners are typically harder to comprehend (Sadighi & Zare, 2006;Tyler 2001). Also of note, was that when learners were given exposure to information relating to the listening text, prior to listening to it, it was found that this resulted in an improvement in topic comprehension (Chang & Read, 2006). Differences were also found in comprehension, especially between academic and non-academic texts, with academic texts seen as more challenging (Ying-Hui, 2006). Significant differences were found in how learners coped with academic and non-academic texts by Buck and Tatsuoka (1998) too, who used the rule-space methodology to attempt to diagnose learner listening problems. Ferris and Tagg (1996), also found that the way academic and non-academic texts were structured, impacted on listening comprehension.

General understanding may depend on background knowledge, which includes topic, structure, schema and culture (Bloomfield et al., 2010). In an experiment, Bartlett (1932, cited in Bloomfield et al., 2010), played a listening passage in the learners L1, but which related to a well-known story. Bartlett changed some details, as well as excluding others. According to him, learners were unable to recall or relate the story as the change did not fit their schemata. Thus, comprehension and understanding depends very much on how well any listening text fits the listeners own background knowledge.

2.6.2.2. Speech rate

Griffiths (1990,1992), found that when playing listening passages, learners seemed to be happy with a listening speed of 1.93 and 2.85 syllables per second, however comprehension became difficult when the speed was adjusted to 3.75 or 5 syllables per second. Higher comprehension was also indicated at 2.85 syllables per second. These were lower level learners. Recent studies seem to indicate that higher level learners are also affected by higher speech rates (Bloomfield, 2010). Rosenhouse, Haik and Kishon-Rabin (2006), who's research subjects were L2 Hebrew learners, found that when a listening passage was played at 3 or 4 syllables per second in the learners L1 (Arabic), comprehension was reduced. Subsequently, when playing an L2 passage at the same speed, comprehension dropped sharply. Generally (Bloomfield, et al., 2010), speed or faster speech rates are seen to be a major issue with learners. In a context where learners were taught exclusively in the L2 (Flowerdew & Miller, 1992), when learners were interviewed about the pace at which the instructor spoke, all but one indicated that this was a concern. Another set of learners who had been keeping journals also expressed their concern about the speed of instruction. In another research project in which learners were allowed to adjust the speed of the listening passage, fourteen out of fifteen decreased the speed of the passage, with one increasing it

(Zhoa, cited in Bloomfield, 2010). So far, the examples given have related just to speed, with no consideration of other factors.

Other issues that seem to impact on perceptions relating to the speed of listening passages, include "reverberated speech" (Moore, Adams, Dagenais & Caffee, 2007), that is speech that may have an echo in it. This was perceived as faster by native speakers compared to passages in which any extra sounds or noise were either removed, or where the listening passage remained the same, despite the speech rate being the same. Griffiths (1990) found that a combination of difficulty and text length impacted on perceptions of speed. Anderson-Hseih and Koehler (1988) found that heavy accented speech was perceived by native speakers as being faster than non-accented speech. L2 listeners also seem to perceive heavy accented speech as being faster (Cheung, 1994). Some listeners prefer a slower rate of speech from non-native speakers who have different language backgrounds (Derwing & Munro, 2001). Note-taking at particular speeds was said to impact on comprehension, specifically 180 words per minute, with no difference at lower speeds of 120 words per minute (Lin, 2006). Speech rate was also examined as a local rather than global factor in research conducted by Buck and Tatsuoka (1998). It was found that the speed of the text required to answer a task question, was different to the speed of text around it. Other research has considered the impact of multiple factors that may impact on speech rate (Brindley & Slatyer, 2002). These include delivery (live vs recorded speech), number of speakers, e.g. monologue or dialogue, task response type, and number of times the audio is played. In their examination of the various factors within a national language proficiency assessment system in Australia, it was suggested that if other variables were kept constant, that a listening speed of 180 words per minute was preferred to another of 200 words per minute for marginal candidates.

According to Bloomfield et al. (2010), studies with L1 listeners where speech processing load was increased (Wingfield, 2000), would in their view indicate that the same

would apply to L2 learners too. When speakers paraphrase or repeat previously stated text, this is considered easier than where the speaker produces new information and where the redundancy is limited (Stine, Wingfield, & Leonard, 1986).

According to Bloomfield (2010), the effect of speech rate seems to be more visible when manipulated in a contrived situation. Thus, in real contexts, there are a variety of factors that impact on the perception of speed.

2.6.2.3. Vocabulary

The impact of vocabulary on listener perception is also discussed. For listening to be successful, an adequate (Nation, 2001) or minimum (Field, 2008) amount of vocabulary is required. Various percentages or numbers of vocabulary required have been suggested. 5,000 of the most frequently used words cover 90-95% of word tokens (Bongers, 1947, cited in Bloomfield, 2010). Hirsch and Nation (1992), suggest a 95% familiarity with words in order to understand the main ideas of a text. Nation (2001) suggested that vocabulary knowledge influences other variables such as fluency in listening, comprehension ability and world knowledge.

Bloomfield et al.(2010) suggest that because there is no measure that relates directly to vocabulary coverage which measures speaking ability, if 5,000 'terms' are known, there is a good chance that the listener will know what has been said. No evidence other than their review of the literature was the basis of this suggestion.

Although the issue of topic interest, general understanding, speech rate and vocabulary are discussed separately, the complexity of listening requires that these factors be seen as interlinked and impacting one another.

We now look at the proposed diagnostic framework as set out by Alderson et al. (2014) and Harding et al. (2015).

2.6.3 A diagnostic process framework

In their article, Alderson, et al. (2014), examine the diagnostic practices and theories of a variety of professions in order to tentatively put forward a theory for diagnostic testing in second and foreign language assessment. Although there is recognition that diagnostic research has already been conducted in the language assessment domain, their view is that the focus is on methods and particularly on standardised proficiency tests.

They examine the processes of diagnosis in fields such as education, at primary and university level; medicine, such as general practice and hospitals, the process used by mechanics to diagnose vehicular problems, and the computing industry, specifically, those involved in IT support (Alderson et al., 2014).

In examining other professions, they note that diagnosis is a normal part of their routine, something that is not necessarily the case in Applied Linguistics as a whole. By examining other professions, there is the possibility of gaining insight into the field, and possibly applying some of the ideas to language testing. Although other fields may also have theorised about diagnosis such as Reiter (1987), in the field of computer science, these may not be useful (Alderson et al., 2014) for the purposes of taking SFL assessment forward. Nonetheless, a useful point put forward by Reiter is that at the start, diagnosis is perhaps a form of speculation as this is a starting point, as we may not be sure which part is "faulty" (Reiter, 1987, p 63). He believes that assumptions are made about faults, which may include the view that other parts are working fine, yet this may not necessarily be the case. Indeed, it is possible that a variety of parts may be the cause of a fault, and thus, as he states, one should not be 'overzealous' in his views when diagnosing. His views seem quite apt for the listening skill, as it is seen as a complex skill, with intertwining processes, which in reality

has not been yet been fully understood and for which there are competing views (Alderson, 2000; Buck, 2001).

Thus, through interviews, and discussions an idea of what diagnosis is or could be is reported. There were varying views about what diagnosis s. For the mechanic it was a case of "trial and error" (Alderson et al., 2014, p.7), whereas for a special needs teacher it was something that came about after having a general idea of a problem and then going deeper and was expressed quantitatively. For some like the neuropsychologist, diagnosis involved standard tools that measured specific problems. A literacy specialist indicated that making adjustments for individuals which may be based on planning was something that was the norm, however, perceptions about what diagnosis was, was that it would be perhaps more methodical and planned. The neuropsychologist indicated that diagnosis was about finding problems or weaknesses, though in language assessment there tends to be talk of strengths and weaknesses. In his article, Lee (2015) seems to be more pro-active in suggesting that ultimately, the purpose of diagnostic testing (DLA) is to find weaknesses and to find remedies for these.

With regards to training for diagnosis, all fields seemed to have some kind of provision, which included in-house training and formal course, with some providing in-depth training. Many people were also involved in finding their own ways of learning more, however, it was also found that experience is considered an important part of being able to diagnose.

The literacy specialist indicated that their focus was more on lower level skills and not on higher level skills such as "understanding or comprehension" (Lee, 2015).

In terms of training, all interviewees asserted the importance of especially formal training, and being able to share experiences with other colleagues, in informal and formal

contexts. It seemed that there was opportunity for people from the same field to get together with others and learn together.

In terms of diagnostic tools, a wide range of tools were said to be available, many of which were free and which were sometime collated in one place in the form of a database. In many case, descriptions were also available of the problem, and even how these were resolved. Many of the resources were available via the internet and to anyone who needed to search for them.

Experience was seen as an important facet of being a good diagnostician. Whereas tools as well as training may be available, it is the experience of having dealt with something before, and perhaps having come across particular problems frequently, that added to personal knowledge and which allowed for decisions to be made about for example, when or when not to diagnose, or how to diagnose. Thus, continuous and regular diagnosis, helps to build knowledge, and thus aids in making a person a better diagnostician, which could include both formal diagnosis, or even holistic (Alderson et al., 2014, p.11). People in the medical field indicated that early career medics rely more on experience as time passes, as opposed to just knowledge. Thus experience is something that needs to be developed in addition to just being able to apply or use diagnostic tools. Thus, a combination of training, access to resources and experience is the ideal combination for diagnostics. At this point in time, this may well be lacking in the second language acquisition field, but more so specifically in diagnostic listening assessment.

The special needs teacher indicated that speaking to learners, and trying to find out their problems is a way to gauge in a general way what problems learners are having, and there may be instances where they volunteer information about common problems they have. This is helpful in the process of diagnosis. However, "self-reporting" (Alderson et al., 2014,

p.13) may not necessarily be reliable, requiring the teacher to look at the students work to confirm or otherwise.

The diagnostic process it seems was found to be primarily listening first, followed by observation, this also involved activation of knowledge and experience and then depending on initial judgement, a decision as to whether a diagnosis is successful, or if further diagnosis or action is required. Alderson et al. (2014) also found that diagnosis can be uncertain, that the individual may not share everything, resulting possibly in 'inaccuracy', it is not necessarily 'black and white' (GP and teacher) , neither is it an 'exact science' (GP), and there may be a need sometimes for "subjective judgement" (Alderson et al., 2014, p.17). The GP alluded to the issue of reliability of judgements, and that there were times where appropriate measures or expertise may perhaps not be available, but that there was no choice but to just 'trust' the information (Alderson et al., 2014, p.17), but there was a need for expertise and experience. There was thus also a need to know one's own limitations and strengths, and to be able to seek further advice or expertise when appropriate. Where uncertainty exists, repeated diagnosis should be conducted. Interestingly, the literacy teacher indicated that normally children were judged on the basis of one diagnosis, however, what was really needed was "constant monitoring" (Alderson et al., 2014, p.17).

Diagnosis and treatment was seen as something separate by those in the medical profession. For example, it was suggested that a diagnosis may occur, but perhaps it may be too late to provide a remedy. I can see this being similar to a situation in the classroom. Some problems may possibly not be treatable, because there are other important goals that relate to a curriculum goal that need to be prioritised, or there may simply be a lack of help available, despite the diagnosis, certainly an issue in the context I work in. However, despite this, there may be an expectation that the teacher needs to devise a solution. The literacy teacher indicated that if a problem cannot be resolved, strategies may be put into place which could

help to manage the problem. This person was talking about a learner with dyslexia, however, in terms of learning, other learning strategies could perhaps be used, though this could still be seen as a remedy.

Overall, it was found that, diagnosis involved identifying problems, and at times solving these, with weaknesses being the target of identification. There are batteries of assessment tools available, and the provision of training, which is greatly improved through experience and the involvement of other stakeholders. Listening seemed to be a common stage at the beginning of any diagnosis, followed by an initial hypothesis, using appropriate tools, resources and own judgements and possibly help from others before making decisions (Alderson et al., 2014). Experience (something that at this point in time is lacking in diagnostic listening assessment) and thus experienced judgement is also seen as very important.

After having examined the practices of other professions, five 'Tentative principles for diagnostic SFL assessment' (Alderson et al., 2014) were proposed.

These principles for diagnosis of strengths and weaknesses are summarised as follows:

1. *It is not the test which diagnoses; it is the user of the test.*
2. *Instruments themselves should be designed to be user-friendly, targeted, discrete and efficient in order to assist the teacher in making a diagnosis. Diagnostic tests should be suitable for administration in a classroom, designed or assembled (with recourse to existing suites of tools) by a trained classroom teacher (or other experienced language teaching professional), and should generate rich and detailed feedback for the test-taker. Most importantly, useful testing instruments need to be designed with a specific diagnostic purpose in mind. This principle is derived from the emphasis the interviewees placed on tools with a clear focus and capacity to play a facilitating role.*
3. *The diagnostic assessment process should include diverse stakeholder views, including learners self-assessments.*
4. *Diagnostic assessment should ideally be embedded within a system that allows for all four diagnostic stages: (1) listening/observation, (2) initial assessment, (3) use of tools, tests, expert help, and (4) decision-making. Much current diagnostic testing arguably begins at stage (3), using general diagnostic tests for whole populations rather than more targeted measures that have been selected on the basis of (1) and (2)….A theory of diagnosis should not preclude large-scale assessments, but it should also pose a challenge to these programmes: would the same decisions about*

strengths and weaknesses have been made on the basis of an individualised assessment in a classroom context?

5. *Diagnostic assessment should relate, if at all possible, to some future treatment.*

(Harding et al., 2015, p.318)

There are implications in terms of research that are connected to the five principles mentioned above (Alderson, et al.,2014).

Teachers require the requisite skills, training and a base of knowledge so that they are able to diagnose, or act as diagnosticians. Research would be required to ascertain how best to achieve this, as it requires appropriate knowledge and a variety of diagnostic tools.

Tools need to be developed that focus on particular points of diagnosis. Reference is made to 'valid' (Alderson et al., 2014, p.22) tools, which is in direct contrast to what was suggested earlier by McKay (2006) about the process of diagnosis (see 2.1.1) as not requiring the kind of validity sought in high-stakes testing. If diagnostic assessment is to progress diagnostic tools need to be developed and made widely available so that the field can be "professionalised" (Alderson et al., 2014, p.22)

Self-assessment needs to be incorporated into the diagnostic process, but at the same time, other more objective tools need to be available to make the data more detailed and insightful. Diagnosis should be integrated into the classroom and linked or fed back into the curriculum. There also needs to be clarity on how the process will work in the daily classroom lesson.

The effectiveness of diagnosis and especially intervention would need to be investigated. Alderson et al.(2014) describe these principles as broad and tentative as the primary source that resulted in these principles are other contexts , which may not necessarily fit into an Applied Linguistics, language assessment context.

For example, whereas cars and human beings have characteristics that are 'normative', language development may not be so clear or linear. They suggest that diagnosis might link to syllabus goals and learning outcomes, or perhaps be based on a theory of SFL

learning. Furthermore, in clarifying the principles, they also caution against somehow making diagnostic assessment, something that becomes mechanical in nature, or perhaps something that starts to dominate. Rather, they believe it should be seen in the context of other facilitating or adding to the repertoire of other classroom assessment tools. In terms of the emphasis on weaknesses in other professions, Alderson et al. refrain from this focus and encourage the use of diagnostic information in conjunction with other assessment that reports on strengths.

In looking at the five principles, principle four is possibly the most important for the authors, as it embodies principles 1,2,3 and 5. Principle 4 is illustrated pictorially in Harding et al. (2015) and is shown below again.

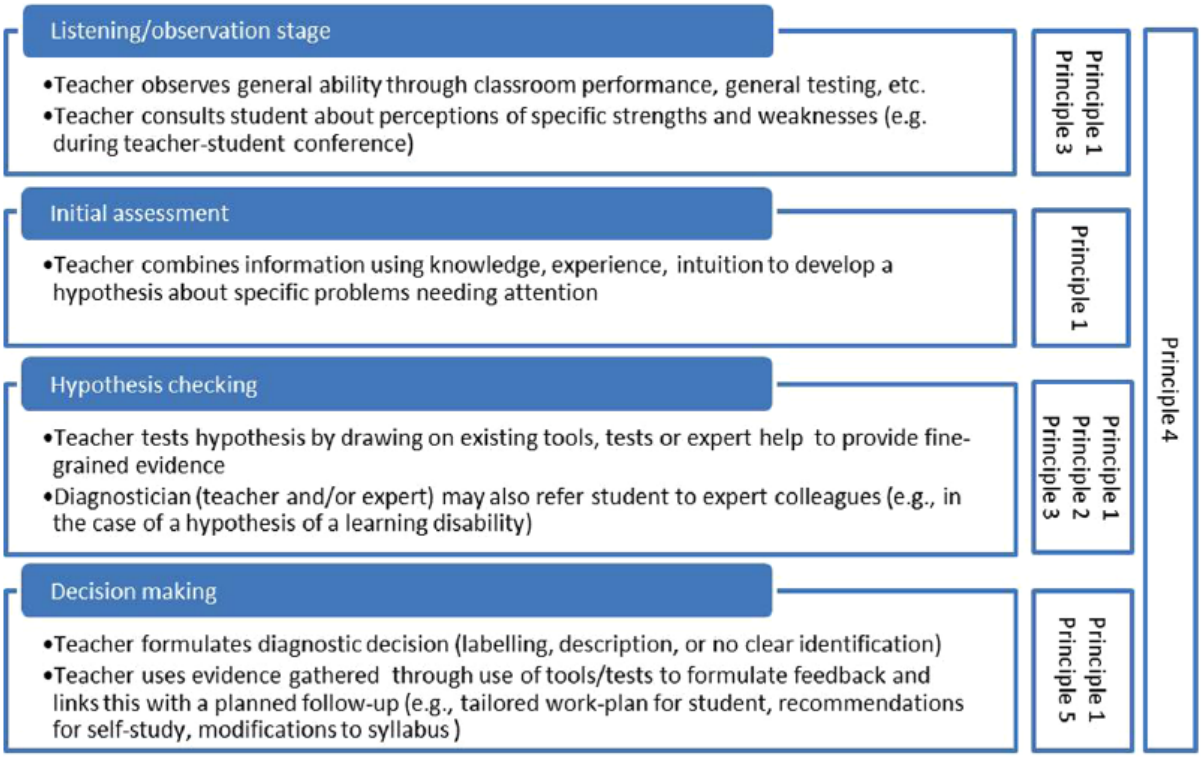


Figure 2 The Diagnositc Process (Harding et al., 2015, p.319)

If there is an element that is perhaps missing from the study conducted by Harding et al. (2015), it is the lack of a context that replicates a typical classroom setting. In classrooms, there is one teacher, and many learners. Classrooms are fluid contexts, where there is

constant change that is not always predictable. Teachers require flexibility and not too much in the way of rigidity. Despite this, it is still worth trying out this diagnostic framework within a typical classroom setting.

This model, according to the authors, has not yet been tested in the language assessment field. (Harding et al., 2015), however, they accept that some of these ideals probably have been applied in writing contexts. In terms of reading or listening however, much less is available in the literature, especially listening assessment in the classroom (Vandergrift & Goh, 2012) where the teaching of listening follows more or less a listening testing model (Field, 2008).

Thus, this research project attempts to investigate diagnostic listening assessment in the classroom, with this framework in mind.

2.7 Summary

This research therefore sets out to ascertain the appropriateness of the diagnostic process as set out in the five principles, which fall within the diagnostic framework set out by Alderson et al. (2014) and Harding et al. (2015).

2.8 Research Questions

The research questions are reiterated below:

RQ1 – To what extent is it feasible to diagnose listening problems in a classroom setting as part of every listening lesson within a Saudi Academic EFL context?

Followed by:

RQ2 – Do the resulting individual and class profiles raise an awareness of learners' strengths and weaknesses in EFL listening from a teacher perspective?

RQ3 – To what extent are learner responses to the listening review sheet (LRS) related to the lexical characteristics of listening passages from an EFL academic text book?

We now look at the next chapter which describes the research methodology.

Chapter 3 RESEARCH DESIGN AND METHODOLOGY

3.1 Overview

This chapter begins by introducing the aims of the study (3.2), the research questions (3.3), the research context and setting (3.4), followed by the research paradigm and design (3.4), the research context and setting (3.5), details of the research participants (3.6), the ethics (3.7), the data collection and research tools (3.8), the piloting of the tools (3.9), the methods of data analysis (3.10), followed by the limitations (3.11) and a summary (3.12).

3.2 Aims of the study

As already stated, this research was motivated by a call for research into the diagnosis and assessment of second and/or foreign languages (Alderson, 2006; Alderson et al., 2014; Harding et al., 2015). The aim of the study was thus to contribute in some way to this call by exploring specifically ways in which listening problems could be diagnosed in the classroom. In order to be able to address this rather new and relatively unexplored area, the research was conducted in an exploratory manner. The issue of whether diagnostic listening assessment can be implemented in the classroom was also an aim, especially how practical (Alderson, et al., 2014) and feasible assessments of this kind could be. With set classroom lessons and well established methods of teaching listening, how would it be possible to collect data that would aid the teacher in becoming more aware of individual listening problems. Thus, how realistic was it to collect information that could help teachers to begin to diagnose listening problems (see section 2.6.2) or at least become aware of the kinds of problems learners face in the listening classroom. What could be done in the classroom that could help or aid teachers to become more aware of, and to begin to diagnose what problems learners were having. The

aim of the study was also to look beyond purely the listening itself, and to try to consider external non-listening factors that may overlap with the problems of listening, for example learner participants may have difficulty in understanding the task, which may contribute to the perception that learners have problems with listening, when in fact task characteristics may prevent them from doing what is required of them. The following section outlines what formed the basis for this research project.

3.3 Research Questions

The main and overarching question is:

RQ1 – *To what extent is it feasible to diagnose listening problems in a classroom setting as part of every listening lesson within a Saudi Academic EFL context?*

The purpose for having this overarching question was firstly, because the research was exploratory in nature. It was also to reflect the thinking behind the overall research, being the need to make diagnostic listening assessment a daily practical undertaking that could potentially impact on the daily listening lesson, teaching and teacher knowledge. Thus, the question tries to focus on finding ways in which listening and task related problems can be identified, which would normally not be the case in a normal listening lesson in which responses to comprehension questions has been the goal (Field, 2008). It is thus the embedding of the diagnosis process and its implementation that underscores this research.

RQ2 – *Do the resulting individual and class profiles raise an awareness of learners' strengths and weaknesses in EFL listening from a teacher perspective?*

Whereas the focus of RQ1 was overarching in nature, this question looked at the usefulness of the data and whether the profiles produced both at classroom level and learner level were in fact beneficial for teachers in terms of being able to differentiate between

learners individual strengths and weaknesses, as well as the broader strengths and weaknesses of the class in terms of specific listening passages.

RQ3 – *To what extent are learner responses to the listening review sheet (LRS) related to the lexical characteristics of listening passages from an EFL academic text book?*

This question took on the role of trying to provide information that could add validity to the data collected in RQ2. Thus, whereas a learner centred, self-reporting classroom tool was used as the central and pivotal data collecting tool during lessons, the purpose behind this question was to ascertain whether there were any links, connections or relationships between what learners perceived about the listening passages, and the lexical characteristics of the listening passages. Thus, it was felt that a lexical analysis could provide a level of additional evidence that could add to the validity of the some aspects of learner views about the listening passages, through the use of the Listening Review Sheet (Appendix B).

3.3.1. Stages of data collection

The table below provides a chronological order of the data collected.

Table 6
Stages of data collection in chronological order

Research tool	Justification	Applicable Research Question (RQ)
Listening Review Sheets (LRS) (Appendix B and 3.8.3)	To collect listening diagnostic information in the classroom. To use the data as a basis for creating profiles of the research participants as a whole, as well as individually	RQ 1,2,3
Phoneme Test	To provide additional information about student listening proficiency	RQ 2
Oxford Placement Test	To ascertain research	RQ 2

(OPT)	participants listening proficiency	
Student Background Questionnaire	To discover patterns if any related to LRS responses	RQ 2
Recorded research participant interviews	To clarify and confirm research participant responses on LRS	RQ 1,2
Analysis of listening passages	To measure lexical and phonological characteristics of the listening passages.	RQ3
Research Questions		
1. To what extent is it feasible to diagnose listening problems in a classroom setting as part of every listening lesson within a Saudi Academic EFL context?		
2. Do the resulting individual and class profiles raise an awareness of learners' strengths and weaknesses in EFL listening from a teacher perspective?		
3. To what extent are learner responses related to the characteristics of the listening passages from an EFL academic text book?		

After research participants signed the consent form, data collection immediately began (RQ1, 2, 3) through the implementation of the Listening Review Sheet (LRS) in the following class. Although the Oxford Placement Test (OPT) (RQ2) should have been the first point of data collection, i.e. before the implementation of the LRS, this still took place in the first week of the research timetable. During the first week, a phoneme test (See 3.8.2) (Appendix U) created by the researcher was also administered. This consisted of dictating words which included letters with sounds not found in the Arabic language and which are considered problematic (Smith, 2001). Once the LRS data collection was complete after approximately five weeks, learners were invited for an interview in order to seek clarification with regards to their responses (RQ 1 & 2). Due to logistical constraints and in order to ensure research participants continued to take part in the research, the student background questionnaire was given to students to fill in just before the interviews (RQ2). It was felt that handing research participants a long questionnaire at the start of the research process, might discourage them from participating. Additionally, in order to sustain research participant

interest, an effort was made to embed the research process into the normal routine of teaching and learning, without seeming like something extra that the learners needed to do. All learners in the class used the LRS (Appendix B), however, only the data from research participants was recorded. The data collected was collated to form class profiles and individual profiles based on the responses to the LRS, and descriptive and inferential statistics produced, using the data produced from the LRS. The listening passages were analysed for lexical qualities (RQ3). I now outline the research paradigm and design in the next section.

3.4 Research paradigm and design

In his article, (Siegel, 2015), Siegel encourages the idea of listening researchers adopting research methods that fit into their contexts, circumstances and whatever tools they wish to use or prefer to use as appropriate. He makes the novel proposal of thinking about 'listening instructors as the participants' (p.3), an idea that has parallels with, according to him, the idea of learners as participants. This new idea is highly relevant to this study as the researcher played a central role in the research project as well as trying to teach at the same time. In effect, Siegel tries to put forward a case for listening research that does not necessarily follow the traditional quasi-experimental or experimental designs as proposed by Cross and Vandergrift (2015), but which is open to any ideas that will help the research process.

In trying to decide on an approach to the research project, a number of factors needed to be considered. The most important factor that helped me in making a decision, was probably the failure to recruit research participants in the first semester of the academic year, and the problem with meeting potential research participants at mutually convenient times outside of the classroom and at mutually convenient times. Problems in the first semester

resulted in delays in moving the research process forward and was becoming a source of concern. Thankfully, in the second semester I shared a particular class with another colleague, who agreed to focus on teaching reading and writing, whilst I took on the responsibility of teaching listening and speaking. This was an ideal situation as I would be allowed to teach and focus on an area linked directly to my research project. It was thus decided that I would research my own class. The research context was thus my classroom, in which the research participants would be a sample of my students.

In addition to this, the necessity and convenience of having to use a sample of my own students as research participants, and I being in effect the main and only researcher, would mean having to conduct the research during normal lessons, and also giving the impression that the lesson was being conducted as normal. The closeness and overlap of the teaching lesson with the researcher and the research, meant that there could potentially be a conflict of interest between delivering lessons as per the teaching schedule, and conducting the research. The research would be very much in my control and I would need to ensure that there were no risks of the research being at risk of any kind of bias.

In effect, looking at the research literature, the situation I found myself in is what one would describe as 'Action Research.' (Kemmis, 1993; Dörnyei, 2007; Cresswell, 2009; Johnson & Christensen, 2012; Cohen, Manion, & Morrison, 2013).

3.4.1 Action Research

Action Research views teachers as researchers, who investigate localised problems, (Johnson & Christensen, 2012), with the goal of trying to improve the practice of teaching, in their local contexts, thereby, taking on the role of building bridges between research, theory and practice. A problem is identified, the literature consulted, and if there is a limitation in the literature, the action research project is seen as a way to solve the particular problem

being addressed (Johnson & Christensen, 2012). During the process of research, the teacher is also seen as someone who reflects on their practice, through a process of organising, trialling, discovering and then contemplation (Kemmis, 1993). The focus is thus the practice that is being researched, with the overriding concern by the teacher researcher to take appropriate and careful steps in a workable, feasible and tangible way, informed by 'practical-theory' (Kemmis, 1993, p.182). Somekh (1995) suggests that Action Research attempts to rectify the problem especially of the 'failure' (p.340) of research to improve practice. Specifically, he suggests that action research does not consider research the first part of a process, and then practice the second. He suggests that in action research, they are in tandem. In other words, that the teacher acts as both, knowing full well the context and being directly involved with the situation being researched (Somekh, 1995). Thus, by way of this process, there is scope and potential to add to the knowledge and understanding of the existing literature upon which the action research takes place (Somekh, 1995).

There are however, some doubts about action research. Dörnyei (2007) believes that it is idealistic as a notion, as in his view, the chances of a teacher researching his or her own practice is unlikely to occur, because he believes that teachers do not have the requisite research knowledge or expertise. It also seems according to him, that even theoreticians or researchers themselves hardly conduct action research, which seems rather hypocritical from his perspective. Additionally, he says that teachers have very limited time, encouragement or grounds to wish to carry out action research.

In their discussion on action research, Johnson and Christensen (2012) dichotomise the role of career researchers and action researchers, by indicating that action research is not really conducted with the view to publishing in academic journals or generalising, but rather dealing with localised particular problems requiring particular solutions. Yet I disagree with this supposition and those by other academic commentators. As someone conducting this

research, I see myself as a bridge builder between theory and practice, and as a professional researcher with a desire to generalise to other contexts the potential findings of this research (See 5.3.2). Conducting the research myself in a practical and concrete teaching context is a necessary step in order to realise the potential of what diagnostic assessment is and to further its development. We now look at the research context.

3.5 Research Context and Setting

The research was conducted within a Saudi Arabian context within one institution. The institution is the English Language Centre (ELC), a foundation year programme that is part of the university, and which straddles between high school and university degree programmes. Once a learner successfully passes the year, he or she progresses to a degree programme. The primary purpose of the programme is to give learners exposure to more English, through a sixteen hour a week English course for one academic year (approximately 400 hours), and to give them exposure to academic subjects that they may have already studied at school, but this time, the content is studied through English. The students where I teach are all male, as in this context, males and females study independently of each other. The students are aged on average between 18 to 22. The context and circumstances were challenging, as learners were very busy with other studies and had very limited time to be available outside class time. We now take a closer look at the research participants. For five weeks, twice a week, learners were handed a LRS. Thus, this became in effect, and integral part of the listening lesson.

3.6 Research Participants

The profile of the research participants was that they were all Saudi, male, and aged from 18-23. There were nine research participants in total which was at the beginning of the

semester about a third of the class. It was deemed appropriate that a sample of the students would participate in the research process as this would aid in restricting the collection of data to manageable levels. Research participants were self-selecting, with no compulsion to be involved. In terms of how representative the sample of students were of the class in general, the results of the Oxford Placement Test indicated that they represented language proficiency levels from sub- CEFR A1 to CEFR C1 (See Table 11). Researching students from my own class had ethical implications which I discuss below.

3.7 Ethics

After discussing the issue of being the sole researcher and who was going to research his own students, my supervisor felt that as long as I was clear on the ethical implications, that this could be done.

In order to ensure I followed proper ethical guidelines, I initially filled out an ethical guidelines form as required by the University of Bristol (Appendix A). Interested learners who came forward were initially involved in a registration process, the data of which was to be kept confidential. This spreadsheet included the names of the participants, as the actual consent form (Appendix G and Appendix H) did not require them to write their names. After finishing the registration process, the researcher asked learners to read a research information sheet, and then sign the consent form. The information sheet and consent form was translated into Arabic (see 8b). The consent form was adapted from a form seen on https://www.nottingham.ac.uk/educationstudentintranet/researchethics/information-sheet-consent-form.aspx

In terms of safe-keeping and securing the data, all of the data that was collected both written and on computer, was kept safely in a file that was not accessible by anyone except

myself. We now take a look at the stages of the research, the research questions, and the research timeline.

3.8 Data Collection Methods and research instruments

A variety of tools and instruments were used in the research study (see Table 3.1 & 3.2 above). Before describing each research tool, the table below illustrates how each tool fits into the research process along with the expected data that was produced. These instruments came into play after the consent form (Appendix G & H) had been signed.

Table 7

Description of Data Collection method, data collected, purpose of data

Ref:	Description of Data Collection Stage	Data collected	Purpose of data
3.8.1	OPT Language Proficiency Test	Numerical score and CEFR Level	Background data
3.8.2	Phoneme Test	Quantitative	Background data
3.8.3	Twice weekly learner self-assessment/Listening Review Sheet	Quantitative Descriptive (Qualitative) *Provides an insight into how the learner is coping with the content of the course*	Class & individual learner profiles Descriptive and inferential statistics
3.8.4	Student Background Questionnaire (SBQ) (Appendix I &Appendix J)	Paper based form	Biodata (Qualitative/ quantitative)
3.8.5	Student Interviews and reflections on using the Listening Review Sheets	Qualitative	Class & individual learner profiles
3.10.3	Lexical analysis of listening passages using a variety of tools including:	Quantitative	Lexical analysis using a variety of established measures & tools

3.8.1 OPT Language Proficiency Test

To get an idea of student's language levels, learners were asked to take an externally validated English proficiency test. It was felt that this would add useful data to the student profile and could potentially be useful for comparing the students when looking more closely at the data when analysing the results. Students were asked to take The Oxford Placement Test (OPT), a widely used online test. The OPT is divided into 2 sections, specifically, *Use of English* and *Listening* (see https://elt.oup.com/catalogue/items/global/exams_testing/9780194571548?cc=fr&selLanguage=en&mode=hub), with the *Use of English* section focussing primarily on Grammar and vocabulary, and the *Listening* section being focussed on listening skills. It was decided to use the listening component CEFR score as a measure rather than the overall CEFR as this was deemed more appropriate for the research. There is very limited writing involved in the test, with no speaking component. According to the website, the test has been pre-tested by more than 19,000 students in over 60 countries.

The results are reported in terms of the Common European Framework (CEFR)(Council of Europe, 2001a); however, the results can be broken down into scores for various sections of the test. It is also available in both American and British English versions, and is entirely online. When allocating the test, test administrators are given the option to select American or British English so that the test items reflect this. As we were using an American series of books, I decided to select 'American' for the Language Use section. For the listening section however, I decided on going for a mix of American and British English. I felt that as I was the teacher, and my accent was in fact British, that students should not be at a great disadvantage and will have normalised to my accent. Thus, listening to both the American and British accent would not have posed a great problem for the research

participants. There is also an option to control the time allocated to the test, and thus I allowed for 2 hours, which seemed more than sufficient.

Students were sent an email link with a username and password. They had to click on the link provided and then do the test. The tests were done on their own laptops at home without any supervision. Although students were sent the link, I had to follow up the emails as some were very slow in doing the online tests. I was tasked with requesting them to do the online test for a few weeks, with some students claiming they had problems with their username and password. I had to in some cases resend the emails. One of the research participants was not successful in being able to do the OPT test.

Interestingly, the CEFR has the bands A1, A2, B1, B2, C1 and C2, with A1 being the lowest band, C2 being the highest. The OPT also has a band called A0 which in effect indicates a sub-A1 score. Thus, learners are deemed to be below beginner level or A1 level.

3.8.2 Phoneme Test

A 28 item phoneme dictation test created by the researcher but not validated was also taken during one of the lessons. The purpose of the test was to add useful additional information to the student profile. Like the Oxford Placement Test which provided a CEFR banding, the phoneme test would provide additional useful information and would focus on known listening problems in the Arab world, specifically with the issue of differentiating between the letters ‘P’ and ‘B’ (Smith, 2001). As the purpose of the research was exploratory, it was felt that such a test would still add value when looking at and analysing student profiles. It also provided the kind of information that is not reported in the CEFR and would also be useful in comparing the CEFR and the phoneme test results as well as be useful in terms of comparison with the diagnostic information produced during the course of the research project. According to Buck (2001), this kind of test succeeds in identifying listening

problems at a very minute level, however, in his view, it is 'unnatural'(p.63) as it provides only phonemic information, and without context. He does however think it is a useful test where the L1 is common, which was the case here.

The students listened to the researcher read out aloud the words which were each read out twice, with intervals of approximately five seconds between the first and second reading. There was a fifteen second pause before reading the next word in the list. Research participants then had to write down what they heard. An attempt was made to read the word list at normal speed. All words except for two began with the letter P or B, with two words ending in B. The limitation of this test was that there was potential to have been more scientific and creative in designing it. For example, words with these consonants at the beginning, middle and end of words, as well as multiple instances of these consonants could also have been included. A full list of the words can be seen in Appendix U.

When marking the results of the test, the spellings of the words were not as important as whether research participants had written the correct letters 'P' or 'B'.

3.8.3 Listening (diagnostic) Review Sheet (LRS)

According to Harding et al. (2015), the diagnostic process should include the views of a variety of stakeholder, one of the most important of which is the learner himself. A self-assessment form (Oscarsson, 1989) or self-reporting questionnaire (Bachman, 2004) allows the researcher to understand what is going on in the mind of the learner, which is appropriate in the context of researching perception skills such as listening. This also raises awareness in learners about the kinds of difficulties they face. Thus, the listening 'diagnostic' review sheet played a significant and important role in the research. Specifically, its purpose was to act as a data collection tool that could potentially help teachers to obtain diagnostic information about individual learners and it was hoped would also raise awareness amongst the learners

about the problems they have. It was hoped that the analysis of each listening review sheet would help the teacher gain insight into what was causing learners to have problems during the listening lesson. The Listening Review Sheet was divided in 24 points that needed to be filled in by the students before, during and after the listening clip was played.

The listening review sheet (LRS), was the most important instrument during this research, and the data collected through this instrument played a pivotal role in the implementation of this research project.

The Listening Review Sheet (LRS) was used by all of learners, however, only the data from the research participants was recorded for research purposes. It became part of each lesson in which listening was covered and the LRS was handed out to students before any listening actually occurred. In each chapter or unit of the book that we used, there were two major listening audio clips which played a central role.

A sample of the Listening Review Sheet is in Appendix B; however, I list the questions and responses that were required in the LRS below. The purpose of each section and question is explained further below.

By the end of the research process, the LRS included the sections: Basic bio-date, Pre-listening section 1, Pre-listening section 2, During listening section and Post-listening section. Here are the questions and responses required:

Basic bio-data

University ID number

Name:

Listening Text name:

Chapter/Unit name and listening text number:

Pre-listening Section 1

Please answer this question in Arabic. Do you know anything about this topic? If yes, could you write as much as you can? Then try to write down what you think you'll hear.

Pre-listening Section 2

Please answer this question in Arabic. What kinds of English words do you think you might hear? Can you guess them? If you don't know the words in English, write them down in Arabic.

During Listening Section

Can you summarise in Arabic the main points of what you heard? As the teacher to play the audio again if required. Write on the other side if required.

Post-listening Section:

Did you understand the general idea?- Yes / No / Some

Do you find the topic interesting? Circle one – Yes / No / Some of it / I'm not sure

How much of the audio did you understand? Put a CROSS (X) on the line that represents how much you felt you understood.

☹----------|----------|----------|----------☺
0% 25% 50% 75% 100%

Pace or speed of recording (Circle one of the choices on the right)

I thought it was very fast,
I thought it was slightly fast,
It was at just the right speed,
It was slightly slow,
It was too slow

Did you hear any words that you have not heard before? - Yes / No / Some

Were there any words that you recognised but could not remember what they meant?- Yes / No / Some

Did you have any other problems? Please tell us here.

Listen for Main Ideas

Circle the best answer. The questions were Easy, Okay, Difficult

I understand what I was supposed to do? YES / NO / Sometimes

Did you get all of the questions right? YES / NO

Were there any other problems that you had? Please explain. For example Accent/ Sound

Listen for Details

Circle the best answer. The questions were Easy, Okay, Difficult

I understand what I was supposed to do? YES / NO / Sometimes

Did you get all of the questions right? YES / NO

Were there any other problems that you had? Please explain. For example Accent/ Sound

3.8.3.1 The process of using the LRS in the classroom

3.8.3.1.1. Importance of the listening review sheet

Each unit/chapter of the book had two major listening sections, known as listening 1, and listening 2. A listening review sheet was used for each of these and was embedded into the listening lesson so that it became part of the lesson and something that the students filled in before, during and after the actual listening. It's main purpose was to act as a major source

of diagnostic data that the teacher (researcher) could look at and make a judgment about individual learners. It was hoped that the data could then be collated for each listening passage, allowing for the potential to obtain a general overview of how the group of research participants performed overall, as well as for the potential to collect data for each individual learner over a period of time, so as to gradually build up a profile at both class level for each individual listening passage and at individual level of what and where problems potentially existed.

3.8.3.1.2. Pre-listening

There were two sections here (Pre-listening section 1 and Pre-listening section 2) which contained questions and where learners had the option to respond in either Arabic or English. Learners could use Arabic, English or a combination of both languages in their responses. Thus, if learners had an idea of the kinds of words they were likely to hear related to the topic, but did not know the English version of the word, they were encouraged to write down the L1 version. Learners could if they wished, also write transliterations of the English words too. It was hoped that this flexibility would encourage learners to delve deep into their minds to search for relevant information related to what they were about to listen to. The primary purpose of this section was thus to get the learners to think about what they were about to hear, to help them predict, and to activate background knowledge if available. It was also an opportunity to ascertain if the research participants were actually interested in the topic concerned. Thus, the more that was written at this stage, would indicate potential for easier comprehension due to familiarity, as well as interest in the topic (Bloomfield et al., 2010). It was also an opportunity to ascertain what learners were thinking, and to give them the freedom to express themselves. Allowing students to write down their thoughts would, it was hoped, give them time to think without the pressure of having to produce immediate

responses to teacher questions. It was hoped this would provide learners the opportunity to demonstrate properly their knowledge level in relation to the topic concerned. They also had the opportunity to predict the kinds of words or vocabulary that they were likely to hear. Questions were both in English and in Arabic.

After these sections were filled in, the teacher (researcher), asked the learners to read the main idea and detail questions in the text book, that came with the listening audio clip in order to 'prime' (Field, 2008) learners towards the listening and to ensure that they were ready to listen and focus.

3.8.3.1.2. During listening

Once the learners had read the questions and tasks in the book, the researcher played the audio clip. During the course of the listening, research participants were required to fill in a section with any points that they remembered. Notes could be in English or Arabic. In order to assist the above, the method adopted for playing the listening clip, was to stop the listening after approximately one minute intervals, making sure to take into account the completion of idea units (Buck, 2001; Field, 2008; Field, 2013). This it was hoped, would reduce the cognitive demands placed on learners as a result of the discourse representation (Field, 2013) created in the mind of the learners to that point. The break would allow learners to write down any points they remembered, and possibly match what they had heard so far, to correct responses in the tasks. The idea of taking breaks was motivated by Browns'(1986) ideas, that like writing, students need to start off with smaller chunks of strings of words, rather lengthy writing. In terms of memory, there was less load required in having to recall what was said after a minute. Although one could argue for the audio clips to be stopped after let's say 15 or 30 seconds, in reality, the ideas in the audio clips needed a certain time to be developed and thus it made limited sense to stop half way through an idea. It was also

convenient from a consistency point of view. Every time the audio was stopped, students were asked to write down anything they could remember on the LRS and also try to answer as many of the main idea and detail questions as possible. This would allow for the notes taken and 'fresh' working memory to be applied immediately to answering questions. The listening passages were relatively long at about 3-5 minutes each (Appendix P) so it made more sense to stop. Sometimes, the recording would be stopped to deal with specific questions that the class had problems with. Stopping the listening once an idea unit was complete, was also consistent with the test process as this had the potential to match an idea unit to a question or item (Field, 2013), thus forming a part of the overall representation as a result of the listening. After students had completed answering the questions and these had been reviewed together in class, students were asked to fill in the remainder of the listening review sheet. and if the students requested, a second playing of the audio would occur, but this time the whole audio clip would be played at once. The second playing of an audio clip was a rare occurrence throughout the research cycle, and was therefore not used as a factor for diagnosis.

3.8.3.1.3. Post-listening

Once all of the tasks in the main idea and detail sections had been completed, research participants were then asked to respond to the questions about the listening passage. The post-listening section covered areas on their general understanding, topic interest, percentage understanding, perceived speed, vocabulary and (added slightly later) success with dealing with the tasks in the main idea and details section.

The percentage understanding was a key question as it was meant to give the teacher (the researcher) an idea of how much was understood. Research participants were presented with the following:

☹----------|----------|----------|----------☺				
0%	25%	50%	75%	100%

The pictorial representation was present on the form to represent the percentage understood based on the research participant's perceptions. The pictorial representation above was adapted from a questionnaire used in another research article (Isaacs, et al., p.40-41, 2015) and is ordinal in nature (Bachman, 2004) . Research participants needed to indicate with a cross on the line the extent to which they understood the listening passage. 0% indicated no understanding whatsoever, whereas 100% indicated full understanding of the listening passage. It was thus meant to indicate to what extent research participants were coping with and understanding the listening content.

3.8.3.1.3. Additional questions added to original listening review sheet

As was mentioned earlier (Section 3.3), the research was iterative in nature, especially in the sense of modifying the LRS as this was the core tool that was to collect the 'diagnostic' data. After having used the listening review sheet for one listening session, it was felt that further questions needed to be added, that could potentially shed light on problems learners may have during the listening task. These questions focussed on problems learners may have with the tasks, or questions. Whereas diagnosing listening problems was an objective, were there non-listening factors that were also contributing to not being able to obtain the correct response?

In terms of the potential for responses from the LRS to map to Field's Cognitive Framework for Listening, the following table illustrates potential points:

Table 8

Listening Review Sheet (LRS) potential mapping to Field's Cognitive Listening Framework

Ques		Response type	Which information could map to Fields Cognitive Processing Listening framework?
	Pre-listening		
1	Do you know anything about this topic?	Text	√
1	If yes, write as much as you can	Text	N/A
1	Write down what you think you'll hear	Text	N/A
2	What kind of words do you think you'll hear? Guess	Text	N/A
	Listening		
3	Summarise, or write down what you heard	Text	√
4	Did you understand the general idea?	Y/N	√
5	Did you find the topic interesting?	Y/N/Some of it/ I'm not sure	√
6	How much of the audio did you understand?	Put a (X) for %age	√
7	Did you hear any words that you have not heard before?	Y/N	√
8	Any words recognised not remembered	Y/N	√
9	Did you have any other problems?		√ e.g. accent – INPUT decoding
	Post-listening		
10	Main Idea/details – Questions/tasks	Easy, okay, difficult	N/A
11	I understood what I was supposed to do	Yes, No, Sometimes	N/A
12	I understood all the words in the questions/answers	Yes, No, Sometimes	N/A
13	Did you get all the answers correct?	Yes, No	N/A
14	Were there any other problems that you had? E.g. Sound, accent, anything else		√ INPUT decoding perhaps

3.8.4 Student Background Questionnaire (SBQ)

After about five weeks of data collection in the classroom, students were invited for an interview. However, before the interview began, students were asked to fill in the Student Background Questionnaire (SBQ) (See Appendix I & Appendix J). It may seem odd that the SBQ was filled in nearer the end of the research data collection; however, I felt that student time was at a premium, and to ask them to visit me and fill in the SBQ, and then ask them for a second meeting to be interviewed may be too much for them. As mentioned earlier, after the earlier problem of research subject recruitment, I did not wish to place too much of a

burden on the students, or disrupt their normal routine. I therefore believed that it was appropriate to ask them to fill in the SBQ at this later stage, as the information that would be written on the form would be unlikely to change over the course of a 4 to 5 week period. The SBQ had originally been created in English, and then translated into Arabic. Each research participant who came for the interview (see section 3.7.2.5 below), was asked to fill in a student background questionnaire (SBQ).

3.8.5 Learner interviews - post listening review data collection

After having spent just over a month using the listening review sheets in the classroom, students were asked to come for interviews to review their listening and for them and the researcher to examine together the listening review sheets.

The interviews were used to check student responses to the LRS, and thus the questions for the interview process followed the order of the LRS questions, and used the exact wording. However, I wanted the flexibility to start with questions where there was possibly information missing, or where the response was perhaps unusual. Furthermore, I wanted to have the flexibility to ask open ended questions sometimes too. In this sense, the interview was guided by the questions on the LRS, yet I opted not to always follow the order of the questions. Thus, although the questions were in order and in a structured manner, there needed to be sufficient flexibility in order to explore any points made during the interview that seemed to be worthy of further investigation. This method of interviewing has been described as 'semi-structured' (Dörnyei, 2007).

The student listening review sheet (LRS) data had already been input into excel spread sheets except for the sections where there was writing involved as this varied and was unstructured, whereas Microsoft Excel seemed more conducive to discrete items of information, thus, only items on the listening review sheet where a response had to be

selected, were recorded. It was felt that the listening review sheets could be physically present and referred to. This was often the case, where a review sheet would be placed in front of the learner, and questions asked about why they chose the responses they did. At first sight, it may seem odd that interviews were taking place so late after the event, but lack of learner time and other commitment by both research subject and researcher, and the constraints of the system did not allow for regular weekly meetings. Thus, this was a limitation of this part of the process. During the interviews, I reminded learners of the listening passage and the topic discussed and opened the book page which contained a short intro and the questions. This helped to jog the memories of the learners. However, once they saw their own writing on the listening review sheets, they were able to give ideas as to why they chose certain responses. It has to be accepted though, that something like a verbal recall protocol immediately after the teaching event, would have been ideal, but the situation did not make this feasible. Furthermore, as the purpose of the research was to look at the feasibility of including a diagnostic process within the listening lesson, or where meeting learners within the constraints of teacher commitments, research protocols of this nature were not deemed suitable. In this sense, a limitation of what can be achieved with such a tool was already evident. Interviews lasted up to 45 minutes and involved going through each listening review sheet and asking questions about and clarifying reasons for the choices made. In many cases, learners had been absent for some lessons, thus interviews were based on what they had filled in individually. In one case, an advanced learner was present for the interviews to help in understanding where the learner could not express himself clearly in English. Additionally, the researcher's level of Arabic was not proficient enough to conduct interviews in the research participants L1. All interviews were recorded.

3.9 Piloting and checking of research instruments

Although the instruments were originally created in English, and then translated into Arabic, it was imperative that bilingual teachers of English went through them again. Whereas a translation may be precise, the question of whether this makes the document suitable for learners to complete is another matter. Therefore, teacher expertise was required, in order to ensure that the instruments were easily understood by the research participants.

The forms that required translating and piloting were the: Research consent form – Students (Appendix G & Appendix H) and Student Background Questionnaire (SBQ) (Appendix I & Appendix J). Piloting included going over the Arabic translations of forms that were to be filled in by the research participants (the learners) by a group of expert teachers. Once a group of expert teachers went through the documents, a group of students similar to the target research participants were asked to go through the forms.

Three expert teachers were recruited for the purpose. Before checking and piloting the instruments, the teacher experts were asked to fill in a teacher expert form known as the 'Piloting of research instruments by experts' form (see Appendix K). The table below summarises the main information relating to the expert teachers:

Table 9

Teacher Experts

Teacher code	Qualifications	Teaching Experience (Years)	Proficiency in Arabic (self-rated)	Proficiency in English (self-rated)
OS-T	BA (Arabic Language) MSc (Tesol)	20 +	Very Good	Excellent
AMS-E	CELTA BA (English & Education) MA (Applied Linguistics & Tesol)	14	Excellent	Excellent
NH-P	BA English Language and Literature	5	Excellent	Excellent

As can be seen from the table, all of the teacher experts had a good level of language proficiency in both Arabic and English. This was required, as it was necessary to look closely

at the aims of the original forms in English, and then to ascertain whether the translations were appropriate.

When examining the Research consent form, the group of expert teachers brought up issues primarily relating to the way the translation came across. Once amendments were made the experts were then asked to fill in the forms as if they were students, and the researcher noted down the amount of time it took to complete this process as suggested by Johnsen and Christensen (2012). All review sessions were recorded to ensure that if necessary, any items agreed could be re-checked.

Going through the student background questionnaire (Appendix I & Appendix J) followed a similar process to the research consent form. Concerns were raised about the kinds of questions formats and responses Saudi learners were used to, and thus the formatting of some questions were changed and where necessary, modified. This was reflected in both English and Arabic versions of the questionnaire.

Once the piloting and checking had been completed by the teacher experts, students similar to the target population were asked to fill in both the research consent form and information sheet, as well as the Student background questionnaire. As this took place during the 1st semester, students from my class were asked to participate on a voluntary basis, to fill in the forms. Students were asked to fill in the forms and mark out with a pen any questions that they thought were not clear, and if they thought appropriate, suggest alternative wording. The time taken to fill in the forms were also timed. All suggestions were marked on the forms themselves, and then an expert teacher was again asked for an opinion on whether any change was needed.

3.9.1. Piloting listening review sheet

Whereas the research consent form and Student background questionnaire had been developed and finalised in the first semester, the listening review sheet went through a few prototypes and then a final design was decided on. Initially an English version was created and then reviewed with teaching expert NH-P. The design was discussed and I explained what kind of information I was trying to elicit. After discussions and based on my own thoughts and reflections, I came up with a final design which was translated by NH-P. Once this was done, which was in the week before the second semester began, I then piloted the listening review sheet with some students from my new class. The process was more about making sure that students understood what was required of them, so in this sense, it was more a process of ensuring that the instructions and the language was clear. No changes were made and the listening review sheet was then incorporated into the classroom.

3.9.2. Amendments to the listening review sheet

After having used the listening review sheet once, I immediately decided to add two more sections. These were, the 'Listening for main ideas' section, and the 'Listening for details' section. This division reflected the way the tasks were divided in the text for all units. In fact every book I have ever used when teaching listening, has divided the post-listening tasks into listening for main ideas and listening for details. By adding these sections, the LRS reflected the whole process of the listening lesson including the tasks. It was believed that this would provide useful additional diagnostic information as the LRS would encompass all elements of the teaching listening process. The importance of adding these two sections reflects Field's (2008) view, that sometimes a great deal of reading or writing may be involved in tasks to demonstrate comprehension. Thus, if a wrong answer is provided, it may not have anything to do with listening at all; rather it may be due to a reading or writing

problem. Thus, learner responses to the questions in these sections could potentially act as a starting-point to undertake further diagnosis and investigation.. This aspect was not piloted, and could thus be considered a limitation in the piloting process.

3.10 Data Analysis

There were three primary sources of data for the research. These were the LRS, the learner interviews, and the listening passages. The table below summarises the resulting output and then how it was analysed.

Table 10

Output of Data Sources

Output of Data Sources

Data source	Output	Method of Analysis
Listening Review sheet (LRS)	Class & Learner profiles	Collated data from LRS Researcher commentary based on visual comparison & contrast of data on spreadsheet Mapping to Field's Cognitive Framework Descriptive statistics Inferential statistics
Interviews	Voice Recordings	Interviews to clarify responses given by research participants Selection of specific important quotes
Listening passages	Lexical and phonological characteristics	Linguistic complexity

As in any research endeavour, there are circumstances which result in differences in terms of the data collected for each research subject.

The table below provides a summary of data collected for each research participant.

Table 11

Summary of data collected for each research participant

Research Participant Code	Completed Research Consent form	Completed Student Questionnaire	CEFR Listening Score	Phoneme Test Score ?/28	LRS out of 10	Interviewed
MOAL01	√	√	N/A	N/A	9	Y
ABAL01	√	√	B1	26	5	Y
SAAL01	√	√	A0 (High)	20	9	Y
YUGE01	√	√	A0 (High)	26	3	Y
OMAL01	√	×	A1	16	3	N
ABAL02	√	√	C1	17	9	Y
AHMO01	√	×	B1	21	6	N
RAAL01	√	×	B1	19	5	N
MOKH01	√	√	A2	21	5	Y

Note: For CEFR listening descriptors from the OPT see Appendix O

There were instances where the listening review sheet was not always properly filled in, resulting in some missing information. Nonetheless, this did not necessarily impede on the overall data and the possibility of making judgements about the data presented.

In total, 10 listening passages were covered in class between the periods 8th February 2016 and 24th March 2016.

3.10.1 Learner listening Review sheets

These played a pivotal role in the data collection with each listening review sheet containing 21 pieces of information, although not all of the data was used. The data produced resulted in a class profile, and an individual profile.

3.10.1.1 Class Profiles

The class profile was created by collating all learner responses for each listening lesson. Thus, collecting all learner responses together for that class would then result in a class profile. This allowed for comparisons and contrasts between learner responses, as well as provide an overview of how the class overall had coped with a particular listening passage. This was done for all listening lessons resulting finally in 10 class profiles.

For each class profile, the following procedure was adopted:

1) The Class profile was split into three sections. The three sections were:

   i) Listening measures, ii) Main Ideas and iii) Details. We take a closer look at them below:

   a. Listening Measures: A summary table based on the first section of the LRS. This included information categorised into the following:

      i. Understand General Ideas
      ii. Interesting
      iii. % age understood
      iv. Speed of Recording
      v. Any words not known
      vi. Words known meaning forgotten
      vii. Any other problems

   b. Main ideas section which resulted in the following data:

i. Main idea difficulty
ii. Understood what to do?
iii. Understood all task vocabulary?
iv. All Questions right?
v. Any other problems?

c. Details section which resulted in the following data:
i. Details difficulty
ii. Understood what to do?
iii. Understood all task vocabulary?
iv. All Questions right?
v. Any other problems?

Where the responses indicated a weakness these were shaded grey. For example, if a research participant responded ‘no’ to ‘understand general ideas, or if he indicated that the ‘recording was slightly fast’ for speed, these would be shaded grey. This would indicate that the learner was having problems that merited attention.

Initially, responses by the research participants for a particular days listening passage were collated in excel. Putting all of the learners data together in this way immediately gave a general overview of how the class viewed the relevant listening passage and demonstrated what a teacher who collected this data would be able to judge about learner problems at a quick glance. Having said this, in a classroom context, teachers may just look through each LRS individually, however, the data becomes more useful when collated together. It was hoped that a general idea or picture could be developed about how the learners were coping, both with the listening audio and the accompanying tasks and where in particular (according to students) they were having problems. A class profile for each listening passage is followed

by description of the data as it would be viewed by the teacher, and additional information from learner interviews, the purpose of which was to clarify and provide further insight into their responses.

Each listening class thus resulted in a class profile. The resulting profile could then be used to gain insight into specific problems, as we all as to get a general overview. An example of the raw class profile data can be seen (Appendix N) as well as in the tables in 4.2.2, illustrating the data for one particular listening passage, with each line indicating a learners responses to the LRS.

As a result of the large amount of data produced, which accompanied each class profile, it was decided to present the full data for only four class profiles in the results section based on the number of research participants. Class profiles 1 (4.2.1) and 4 (4.2.4) had 8 research participants, and class profiles 2 (4.2.2) and 3 (4.2.3) had 9 research participants respectively. Relevant data from the other class profiles that were found to enrich the data were also included.

3.10.1.2 Individual profiles

Learner data for each listening review sheet, was collated for each student, resulting in a unique diagnostic learner or individual profile. Thus, the teacher had at his disposal a unique record of how each learner had reacted to a variety of listening passages. This was useful because it showed how learners coped with a wide variety of listening passages and provided an opportunity to look at general trends in the kinds of problems faced. There was also an attempt to then map some of the data in terms of Field's Cognitive Framework. Individual profiles are presented from 4.3 onwards and follow the same format of teacher commentary, followed by relevant excerpts and descriptions from interviews.

3.10.1.3 Descriptive and inferential statistics

In order to analyse the data from the LRS, using descriptive and inferential statistics, categories in the LRS that were ordinal in nature were converted into numbers. The table below summarises the information that was converted to aid the process:

Table 12

Ordinal Values converted to numerical

Value		Label
Passage	1	BK2U6L1
	2	BK2U6L2
	3	BK2U7L1
	4	BK2U7L2
	5	BK2U8L1
	6	BK2U8L2
	7	BK2U9L1
	8	BK2U9L2
	9	BK3U1L1
	10	BK3U1L2
Understood General Idea	1	No
	2	Yes
Interesting Topic	1	Not Sure 1
	2	No 2
	3	Some 3
	4	Yes 4
Speed	1	Very slow (Score 1)
	2	Slightly slow (Score 2)
	3	Just right (Score 3)
	4	Slightly fast (Score 4)
	5	Very fast (Score 5)
Words Not known	1	All words known
	2	New Words
Meaning forgotten	1	Vocab Meanings NOT Forgotten
	2	Vocab Meanings Forgotten
Main Idea Questions	1	Hard
	2	Okay
	3	Easy
Main Idea Task Understood	1	No
	2	Some

	3	Yes
Main Idea Task Vocab Understood	1	No
	2	Some
	3	Yes
Main Idea All Questions Correct	1	No
	2	Yes
Listen For Detail Questions	1	Hard
	2	Okay
	3	Easy
Details Task Understood	1	No
	2	Some
	3	Yes
Details Task Vocab Understood	1	No
	2	Some
	3	Yes
Details All Questions Correct	1	No
	2	Yes

In addition to the above, the CEFR bandings were represented as follows:

Table 13

CEFR Band Numeric Value

CEFR Band	Numeric Value
Not applicable (N/A)	0
A0 (high)	-1
A1	1
A2	2
B1	3
B2	N/A
C1	5
C2	N/A

The CEFR bandings reflected the OPT listening (see 3.8.1) scores. N/A indicates that no one scored within that band range. There was no calculation involving the CEFR numerical values, which were used primarily to help present data. SPSS version 23 was used to calculate descriptive and inferential statistics.

3.10.2 Learner interviews

All learner interviews were recorded. The recordings were conducted in order to allow learners to explain and clarify why they selected particular responses in the listening review sheets and to ascertain what they believed these responses reflected in terms of their own listening. Where the interview shed light on the reasons for their responses, these were highlighted and used to describe the listening experiences of the learners. As only specific and relevant quotes were selected in order to provide additional description to the results from the LRS, no transcription of the audio recordings were made.

3.10.3 Analysis of listening passages

In addition to analysing the data from the LRS, it was felt that analysing the listening passages would add value to the research project, specifically in that this may confirm or be consistent with learner perceptions of the listening passages. In order to analyse the data of the listening passages, a variety of characteristics were measured.

3.10.3.1. Speed of delivery

How learners perceived the speed of the listening passage was an important measure as this provided immediate information to the instructor about how learners were likely to react to any listening passage. This was measured manually.

A words per minute (WPM) calculation was based on a playing of the audio, which was stopped after the end of each minute. The transcript was followed whilst playing the audio. At the end of the minute, the point at which the transcript had been reached was considered a cut-off point. Words that overlapped the minute point were included in the previous minute. Words were then counted three times, the total then being written down. A

word per second measure was calculated by dividing the words in that minute, by sixty (being 60 seconds). If at the end, the passage ended in between minutes, the number of words were converted into a WPM measure. No additional adjustments were made, for example for pauses. Although software such as Praat was available to measure audio passages, my own limitations in being able to use the software prevented this.

3.10.3.2. Length

The length of the passages was manually counted. Thus, the length was represented by the number of words, as well as a total length in terms of time which was noted from the windows audio player. A table showing manual calculations of the listening property passages can be seen in Appendix P.

3.10.3.3. Lexical Complexity

An analysis of the texts of the listening passages was conducted using a variety of online software tools. It was felt that this information could be useful in providing additional data to possibly explain how the research participants had perceived the listening passages. To aid the process, a variety of lexical tools were used. These were:

3.10.3.4. VocabProfile

This is an automated web based software tool that analyses the lexical qualities of a text. According to the website, (www.lextutor.ca/vp/research.htm), it analyses passages in four ways:

i) Percentage of the most frequent 1000 English words (K1)

ii) Percentage of the most frequent 1-2,000 words of English (K2)

iii) The academic words of English, based on an Academic word list of 570 word families (Coxhead, 2000) that occur across a variety of academic subjects,

iv) Words not found on the other lists

In addition to this information, VocabProfile also provided information on:

3.10.3.5. Type-token ratios

Word 'types' are occurrences of unique words, whereas 'tokens' are the total number of words occurring in the text or passage. The number of unique types divided by the number of words in total (the tokens), results in the type-token ratio (Graesser, McNamara, & Kulikowich, 2011).

Although other data was available from VocabProfile, it was decided to keep data analysis to a specific number of measures in order to keep the information concise.

3.10.3.6. Coh-Metrix

This online lexical analysis tool (McNamara, Louwerse, Cai, & Graesser, 2005), was also used to provide additional information related to the listening text. The additional measures were MTLD, voc-d and concreteness.

3.10.3.7. MTLD and voc-d

MTLD and voc-d which are measures considered better in reflecting lexical diversity compared to the type-token ratio mentioned above (MacWhinney, 2000; McCarthy and Jarvis, 2007; McCarthy & Jarvis 2010). Specifically, the view is that type-token as a measure of lexical diversity is influenced by text length.

3.10.3.8. Concreteness

This measures the concreteness or abstractness of a passage (McNamara, Crossley, & McCarthy, 2011; Révész & Brunfaut, 2013). Though primarily for reading passages, the information nonetheless could provide useful additional information about the passages and its impact on listening comprehension. A higher score indicates greater concreteness, with the opposite being indicative of less concreteness and thus more abstractness, suggesting that the lower the number, the more difficult the text (Graesser et al., 2011).

3.10.3.9. Text Inspector

Text inspector (http://www.textinspector.com/) also analyses texts for lexical diversity, based on the CEFR and The English Profile project (http://www.englishprofile.org) at Cambridge which provides the underlying data which is based on Cambridge Esol exams.

Additionally, useful measures such as the '%age of words with more than 2 syllables', and average syllables per sentence, were also used to analyse the text.

With a wide variety of lexical tools available as illustrated above, and the research being of an exploratory nature, each tool provided additional unique measures that were not always present or easily accessible in other tools (see Appendix X, Y & Z). Therefore, accessing these tools provided an opportunity to select measures that were of relevance or potential relevance to the current research.

3.11 Limitations

3.11.1 Student recruitment/availability

One of the biggest issues was student availability and their commitments towards other academic subjects. As a result, trying to recruit students was not an easy task.

3.11.2. Time constraints

Due to time constraints, limitations were placed on what could be achieved in the research project. Thus, whereas there was a desire to experiment with, and try out possible methods of resolving issues that learners faced in the classroom, time did not allow for this. As a result, the listening review sheet (LRS) became the central aspect to the research that involved learners, and so the focus turned to a tool that provided information that had the potential to be used for furthering the scope of diagnosing listening problems, in a direct and indirect method. Furthermore, the actual answers that learners gave for the listening tasks were not analysed due to time constraints and issues of tool design which would need modifying for each task type.

The time lapse between using the LRS and the actual interviews may well have impacted on the interview process which took place approximately a month after the first LRS was used. Therefore, in an ideal situation, the time between use of the LRS and a follow-up interview should have been much shorter, possibly within the same week. Where students were absent, this meant that they had not filled in a LRS, and thus this reduced the amount of data available for comparison purposes. Furthermore, some students did not respond sometimes to all of the questions in the LRS. Having discovered this, I then provided more class time for filling in the LRS as well as making sure that every part of the sheet was filled in before this was handed to me at the end of the class. Although notes were made that I could use to reflect on, these were not contained in one document or notebook, but using a variety of methods. Having one central point where notes are collected ensures that everything is easily accessible.

3.11.3. Lexical Analysis

Although the LRS contained questions about whether the task prompts and task vocabulary had been understood, time constraints meant that an analysis of the listening passages had to take precedence over a lexical analysis of the questions prompts and responses. A lexical analysis of the tasks could have provided further information that could possibly provide additional information about learner responses that were specific to whether task question vocabulary had been understood.

3.12 Summary

This chapter has presented the methodology and design of all aspects of the research in this research project, as well as describing the research tools, their purpose, the method of data collection and how the data is to be analysed. The table below reiterates the research design and procedures

Table 14

Data collection timeline and procedures

Data collection timeline and procedures
Research Questions RQ1 – To what extent is it feasible to diagnose listening problems in a classroom setting as part of every listening lesson within a Saudi Academic EFL context? RQ2 – Do the resulting individual and class profiles raise an awareness of learners' strengths and weaknesses in EFL listening from a teacher perspective? RQ3 – To what extent are learner responses to the listening review sheet (LRS) related to the lexical characteristics of listening passages from an EFL academic text book?
Strategy Action research
Data Collection Methods • Learners take OPT (Oxford Placement Test) (RQ2) • Use of listening review sheet (LRS) to collect diagnostic listening data based on two listening passages per week for five weeks. (RQ1, RQ2, RQ3) • Interviews of some of the research participants after a period of approximately 5 weeks to gain further insight into LRS responses. (RQ1, RQ2, RQ3) • Student Background questionnaires for those interviewed. (RQ2)

- Lexical analysis of listening passages (RQ3)

Sample:
9 research participants

Timeline:

Data collection:	February 2016 to March 2016
Analysis and writing:	April 2016 to May 2017

Chapter 4 RESULTS

4.1 Overview

As has already been stated, the purpose of this study was to ascertain if a certain set of listening problems could be diagnosed during a typical listening lesson. The data was thus analysed into class profiles (4.2) and individual profiles (4.3).

Section 4.2 presents the data relating to class profiles, which is the data extracted from the LRS filled in by each learner for a particular passage. The results of three passages are presented in sections 4.2.1 to 4.2.3.

Section 4.3 presents data in the form of individual learner profiles. Here, a particular learner's responses for all of the LRS's filled in are combined to form an individual profile. These are found in sections 4.3.1 to 4.3.3

Finally, the lexical qualities of the passages are presented and then compared to research participants perceptions of the listening passages, based on their responses to the LRS. These are found in 4.5.3

4.2 Class profiles

As already stated in 3.10.1.1, the class profiles are divided into three parts based on, 1) learner listening perception measures, 2) Main idea tasks, and 3) Details tasks, and are followed by a commentary and relevant student interview data (see 3.8.5 for details) to help clarify and gain further insight into their responses. These are presented below.

4.2.1 Class Profile 1 - BK2 U6 L1 - Howtoons

Table 15

*Class Profile 1, **LISTENING PERCEPTION MEASURES**, Lesson 1, Book 2, Unit 6, Listening 1 – Howtoons - BK2 U6 L1*

Research Subject	CEFR Listening Score	Understand General idea	Interesting	%age understood	Speed of Recording	Any words not known	Words Known Meaning Forgotten	Any other problems
MOAL01	N/A	Y	SOME	50	3	Y	Y	-
SAAL01	A0 (High)	Y	NOT SURE	35-45	5	Y	Y	-
YUGE01	A0 (High)	-	SOME	25	4	Y	Y	-
OMAL01	A1	Y	SOME	25	3	Y	Y	-
ABAL02	C1	Y	Y	85-95	3	Y	N	-
AHMO01	B1	Y	Y	100	3	N	N	-
RAAL01	B1	Y	SOME	100	3	N	N	-
MOKH01	A2	Y	SOME	60-75	4	Y	Y	-

Speed of recording (speech rate) key:

1= I thought it was very slow, 2= I thought if was slightly slow, 3= It was just at the right speed,

4= I thought it was slightly fast, 5= I thought it was very fast

Speech rate per second: 1.90-2.13 words per sec. Avg.: 2.00

No. of words: 330

Length in minutes: 2.46

For the first listening passage, that is, book 2, Unit 6, listening passage 1, as can be seen, there were 8 research participants present.

Understand General Idea

Of these, everyone understood the general idea. One person did not respond to this question.

Interesting

Two out of eight found the topic interesting, with five finding the topic somewhat interesting. One individual was not sure about how he felt.

%age Understood

With regards to the percentage understood, two found they could understand everything (100%), one person understood from 85-95%, MOKH01 could understand 60-75%. Two it seems understood only 25%, with one understanding from 35-45%, and MOAL01, saying he understood only 50%. Interestingly, looking at the CEFR listening results, those who scored B1 on the OPT listening section, claimed they understood everything, whereas ABAL02 who scored C1 in the OPT claimed he only understood 85-95%.

Those who scored A1 or less, as well as MOAL01 who did not have a CEFR score, seemed to understand at most only 50%.

Speed of Recording

In terms of speed, five found the speed of the recording at just the right speed, whilst YUGE01 and MOKH01 (CEFR A0 and CEFR A2 respectively) found it slightly fast, with SAAL01 (CEFR: A0 (high)) finding it very fast. OMAL01 (CEFR A1) indicated that he found the speed just at the right pace. The speed of the recording or speech rate (Field, 2013:118) varied from 1.90 words per second per minute to a maximum of 2.13 words per second per minute, with an average of 2 words per second (See Appendix P). These speech rates fall at the lower end of the mean speech rate for the KET exams as analysed by Field (Field, 2013:118) which ranged from 2.05 to 2.69 words per second.

Any words not known and meanings forgotten

With regards to words not known or new vocabulary, five found words they did not know, whereas three claimed they had no problem with the any vocabulary including MOKH01 who then went on to indicate that there were words he heard, but for which he had

forgotten the meanings. Thus he was familiar with all of the new vocabulary, but still picked up words that he knew from before. No other problems were indicated relating to these areas of listening.

Table 16

*Class Profile 1, **MAIN IDEAS & DETAILS SECTION**, Lesson 1, Book 2, Unit 6, Listening 1 – Howtoons BK2U6L1*

Research Subject	CEFR Listening Score	MAIN IDEA **TASK** Difficulty	LISTEN FOR Details **TASK** Difficulty
MOAL01	N/A	EASY	EASY
SAAL01	A0 (High)	HARD	HARD
YUGE01	A0 (High)	HARD	MIDDLE
OMAL01	A1	HARD	-
ABAL02	C1	EASY	EASY
AHMO01	B1	EASY	EASY
RAAL01	B1	EASY	EASY
MOKH01	A2	EASY	EASY

Main Idea task: Select 3 main ideas from 6 options

Details task: Select a word for sentence completion

Main ideas and details

As can be seen, SAAL01, YUGE01 and AMAL01 indicated difficulties with the main idea and detail sections.

During a walk around the class, whilst the researcher glanced at learner LRS responses, learners were asked about why they had difficulties. SAAL01 (A0High) indicated that he was trying to develop – "I am trying to develop but no found develop". It was not clear what he meant, but he may have been referring to his English language development.

YUGE01 (A0High) said that he had problems with the speed and the vocabulary. Though it was difficult to ascertain which vocabulary he was referring to, he had indicated earlier that he came across new and known vocabulary that he could not comprehend. However, on further questioning especially relating to his response of 'Middle' relating to the 'Listening for Details Section', he said that the information required to answer the questions seemed too close to each other, thus it made it difficult to answer the questions as there seemed to be very little time and space between the questions. This would indicate that the questions may not have been spaced out properly, thus may be an item writing problem, or that the information density of the passage was greater where the information for the correct response was to be found. These comments were noted on the listening review sheets by the researcher. This and other responses resulted in considering whether more questions needed to be added to the listening review sheet so as to capture more information that could help in diagnosing issues that learners were having in relation to the listening lesson. More questions were then subsequently added to the listening review sheet, specifically to the main idea and details sections. OMAL01 also indicated that he had problems with speed and vocabulary.

Interviews[2]

MOAL01 (N/A)

He indicated that he knew a lot about the topic, and he found some of the passage interesting. He only understood 50% of the passage. He said that there were a lot of new words, and that he also heard many words that he knew, but whose meanings he had forgotten. With regards to the latter, he said: *"all the time this problem comes".*

[2] As already mentioned (see 3.8.5), learners were interviewed in English except in one case (SAAL01) who was unable to express himself clearly in English and thus another research participant (ABAL02) agreed to translate during the interview.

In effect, what he was suggesting was that whereas he was familiar with or had come across lexis that he knew, he could not remember the meaning of the words, and this happened often. Perhaps this was a general area that needed to be considered for potential remedial work.

ABAL02 (C1)

During the interview process which took place after all of the listening lessons had been completed, being about a month later, ABAL02 explained that the 'percentage understood' number reflected times when 'the mind wandered off'. As he said, *"sometimes you go away to your mind and the listening, and something go and you didn't listen, or a new word".*

So basically, he seemed to indicate that his focus may have been reduced, resulting in a few instances where he was not fully concentrating on the listening passage. Perhaps this indicated that there were working memory or concentration issues at play. Thinking about other learners, it is possible that others also had instances where their minds also wandered off, yet they may not have considered this when writing down the percentage understood. I did mention to ABAL02 that it is quite normal for people not to be able to focus all the time. During the interview, ABAL02 indicated that he had heard new vocabulary which he was not familiar with, but also heard words he knew but for which he had forgotten the meanings. ABAL02 had the highest CEFR score amongst the research participants, yet he seemed to be conservative about his listening ability.

MOKH01(A2)

By stating that he understood 60-75% he understood the general idea, but not all of he words. MOKH01 misunderstood the question which asked whether he had come across new vocabulary. He had indicated 'No', but in fact he meant yes.

On looking at the transcript, he said that there were many words but he apparently had not heard them. When asked why he thought this was the case, he believed that this was because the speaker spoke quickly. He also thought that some speakers link the words together, by which he indirectly suggested that perhaps he had problems with segmentation or recognising words when they were linked together during the course of the listening.

SAAL01(A0)

In order to bridge any possible communication problems, I asked ABAL02 who was also involved in the research and who had the highest English language proficiency of all of the research participants, to act as a translator if required during the interview. SAAL01 agreed to this. As ABAL02 had already been through this interview process and was familiar with the types of questions to be asked and potential responses, it was hoped that his involvement as intermittent translator would not have any affect on the interview data. Despite this, it could be argued that there may always be a possibility of translated responses not being exactly as stated by the interviewee.

When I mentioned Howtoons and the fact that I was going to ask questions about this passage and others, SAAL01 could not recall anything. I opened the book and showed him the unit. I also showed him his listening review sheets.

In showing his listening review sheets, I asked why he had not written his responses in Arabic. When filling in the listening review sheets, there was an option for research participants to write in either English or Arabic their views about the passage they were going to hear, as well as notes taken whilst listening to the passage, primarily to ascertain their topic knowledge and predictive ability, and what they were able to glean from the listening passage. He had scored CEFR A0 in the OPT exam, and thus was according to this, a low level learner. He said that: "*I need to learning English*". His view was that his classmates wrote in English and thus he also wrote should write in English. Thus, this raised the question

of how learners perceive the LRS and how they perceived themselves when with peers. It is important that learners feel comfortable and at ease when using these tools, thus perhaps this suggested that proper training on how to use the tool was required, which took into account learner doubts.

Regarding interest in the topic, he selected 'NOT SURE'.

He said:

"Most listening I see the letters without the meaning. Topic I don't the meaning".

To clarify, the translator said that SAAL01 heard the topic, but didn't know what they were talking about. His response indicated that perhaps because he was not able to understand the passage, and thus to say that he found the topic interesting, was a difficult question to answer. This was confirmed by what he stated next.

In terms of understanding, he said he understood 35-45%. I asked if he could clarify what this meant for him. He said the speed and vocabulary were the reason he gave that percentage.

He also did not seem to mention any other problems, but he said this was a lie. It seems that he was embarrassed to write anything in front of his classmates, which I accepted. He wanted to fit into the class. I reassured him that this information was confidential and that he did not have to worry about anything. This raised issues about how 'truthful' learners were when responding to the LRS.

4.2.2 Class Profile 2 - BK2 U6 L2 – Sell-it-Yourself

Table 17

Class Profile 2, LISTENING PERCEPTION MEASURES, Lesson 2, Book 2, Unit 6, Lesson 2 – Sell-it-yourself - BK2U6L2

Research Subject	CEFR Listening Score	Understand General idea	Interesting	%age understood	Speed of Recording	Any words not known	Words Known Meaning Forgotten	Any other problems
MOAL01	N/A	N	SOME	25	4	y	y	-
SAAL01	AO (High)	-	N	0	3	Y	Y	-
ABAL01	B1	Y	SOME	100	3	Y	Y	-
ABAL02	C1	Y	Y	85-95	3	Y	Y	-
YUGE01	A0 (High)	N	NOT SURE	10-15	4	Y	Y	
OMAL01	A1	N	-	0	3	Y	Y	vocab
AHMO01	B1	Y	SOME	100	3	Y	N	
RAAL01	B1	Y	N	100	4	Y	N	N
MOKH01	A2	Y	Y	80-90	3	Y	Y	

Speed of recording (speech rate) key:

1= I thought it was very slow, 2= I thought if was slightly slow, 3= It was just at the right speed,

4= I thought it was slightly fast, 5= I thought it was very fast

Speech rate per second: 2.12-2.58 words per sec. Avg.: 2.33

No. of words: 624

Length in minutes: 4.32

Nine research participants were present for this lesson. The listening review sheet was still something that the learners were getting used to.

Understand General Idea

In the table above, for listening 2, it can be seen that five research participants understood the general idea, whilst three didn't, whilst SAAL01 did not indicate what he thought.

Interesting

In terms of interest, there was some interest in three instances, two found the topic interesting, one person was not sure, while two said no. In effect, more than half indicated that they were not interested in the topic concerned.

%age Understood

In terms of percentage understood SAAL01 and OMAL01 indicated they did not understand or comprehend anything and gave 0%. Two others, YUGE01 and MOAL01 suggested that they could only understand 10-15% and 25% respectively. Two were able to comprehend from 80-90% while three could comprehend 100%. This seemed to reflect the wide variation of listening abilities in the class.

Speed of recording

In terms of speech rate/speed, despite indicating 0% comprehension, SAAL01 and OMAL01 still indicated that the speed was fine as did ABAL01, ABAL02,AHMO01 and MOKH01. Perhaps this indicates that where speed is deemed acceptable, that the problem of comprehension at this point lies somewhere else, not with speed. MOAL01 who indicated understanding only 25% of the passage indicated that the speed of recording was slightly fast. Did MOAL01 perhaps have stronger vocabulary knowledge than those for whom speed was not a problem area, and was speed impacting on his own comprehension? He scored A2 on the OPT, which would indicate some advantage. However, this indicated that perhaps speed was sometimes a cause for lack of comprehension, but not always, though at this point this was difficult to ascertain as those who said speed was not an issue, claimed their comprehension was at a zero level. Interestingly, RAAL01 who said he comprehended 100%

also thought the speech rate was slightly fast. This seemed to indicate that despite the perceived faster speech rate, RAAL01 was nonetheless able to follow what was said. The listening passage speech rate (see Appendix P) varied from 2.12 to 2.58 words per second per minute with an overall average of 2.33 words per second. In terms of the overall average, this passage was slightly faster compared to the earlier passage, and in fact in terms of overall average compared to all other passages, was the second fastest (Appendix P).

Any words not known and meanings forgotten

With regards to the vocabulary, all indicated that they came across new vocabulary that they were not familiar with. With the exception of AHMO01 and RAAL01, everyone indicated that they came across or recognised words they knew, but which they had forgotten the meanings of. Whereas the earlier listening passage had one speaker, and was more like an informational lecture, this followed an interview format.

Interviews

ABAL02

ABAL02 (CEFRC1) again, as for the previous listening passage indicated approximately 80-95% understood. In explaining his percentage during the interview process, ABAL02 said:

"Actually, most of the listening may be all I put this. I didn't agree it come to 100".

So interestingly, ABAL02 seemed reluctant to put down 100% understood, as he felt this was not truly accurate, or may be he felt hesitant about doing this. Interestingly, some research (Unaldi, 2014), does seem to indicate that better students tend to underestimate their performance as compared to other learners. Again, like in the previous passage, he explained this by saying that there were situations where his mind wandered off, in addition to coming across new vocabulary.

Table 18

*Class Profile 2, **MAIN IDEAS SECTION** , Lesson 2, Book 2, Unit 6, Listening 2 - Sell-it-Yourself – BK2U6L2*

Research Subject	CEFR Listening Score	MAIN IDEA **TASK** Difficulty	MAIN IDEA **TASK** Any Other Problems
MOAL01	N/A	-	HARD BECAUSE OF NEW WORDS
SAAL01	AO (High)	HARD	BECAUSE QUALITY SOUND
ABAL01	B1	EASY /HARD	-
ABAL02	C1	EASY	-
YUGE01	A0 (High)	HARD	-
OMAL01	A1	HARD	-
AHMO01	B1	EASY	-
RAAL01	B1	EASY	-
MOKH01	A2	HARD	Much info and similar answers

Task: Select 1 paragraph from 3

Listen for main ideas task

Of the nine research participants, four indicated that the main idea task was hard, whereas three indicated it was easy. ABAOL01(B1) indicated both easy and hard. May be he meant it was somewhere in the middle. It wasn't clear what he meant unless he could explain this. Under 'any other problems', MOAL01 said he found the section hard because of a lot of new words. It was difficult to ascertain what he meant. Did he mean the task instruction vocabulary? The interview process would hopefully clarify this. SAAL01 also indicated that the main ideas section was hard, and then added that there were issues with the sound quality. There is nothing surprising about this, as he had earlier said that he understood 0% of the

listening passage. In this sense, if he could not understand the passage, his chances of answering any questions were unlikely. MOKH01 also added a comment saying that he thought there was a lot of information and that he thought the responses were similar. The points raised by the learners made it necessary to consider providing more questions on the listening review sheet so that problem areas could be narrowed down.

Table 19

*Class Profile 2, **DETAILS SECTION** , Lesson 2, Book 2, Unit 6, Listening 2 - Sell-it-Yourself - BK2U6L2*

Research Subject	CEFR Listening Score	LISTEN FOR DETAILS **TASK** Difficulty	LISTEN FOR DETAILS **TASK** Any Other Problems
MOAL01	N/A	-	-
SAAL01	AO (High)	-	-
ABAL01	B1	EASY	N
ABAL02	C1	HARD	-
YUGE01	A0 (High)	-	-
OMAL01	A1	-	-
AHMO01	B1	EASY	-
RAAL01	B1	OKAY	Too long
MOKH01	A2	EASY	-

Task: Write short notes based on question prompts

Listen for details task

For the details section of the task, three research participants said they found the section easy, while one each found it okay and hard respectively. Four did not put down any comments. RAAL01 mentioned that the length of the passage was too long. In fact amongst

all of the passages, this passage was the longest in terms of words, being 624 words long, but about half a minute less than the longest which was 5 minutes four seconds long. I did however endeavour to stop the recording passage the first time it was played, approximately every 1 minute. Interestingly ABAL02C1 found the task hard.

Interviews

On the listening review sheet, ABAL02C1 wrote:

"They ask for numbers and I am not good in remember number and it was close the two number how say mention".

During our interview, and after having read back his note to him, he stated that:

"there were a lot of number, and I always forget number".

On further questioning, I asked him whether he had problems differentiating between numbers, for example, 15 and 50, to which he said no. So as far as he was concerned, there were too many numbers and this probably confused him, resulting in him finding the section harder than expected. As he stated during the interview,

"it's about memory".

There were in fact eight question prompts, out of which only two related directly to numbers, that related to the number of members and participants in online selling and publishing. He may well have got the other responses correct (which I did not confirm at the time of our interview), however, the difficult questions overshadowed those that gave less or no problems. It is possible that the spacing between questions could have made answering the number questions difficult.

4.2.3 Class Profile 3 – BK2U7L2 – The Great Pacific Garbage Patch

Table 20

*Class Profile 3, **LISTENING PERCEPTION MASURES**, Book 2, Unit 7, Listening 2 - The Great Pacific Garbage Patch-BK2U7L2*

Research Subject	CEFR Listening Score	Understand General idea	Interesting	%age understood	Speed of Recording	Any words not known	Words Known Meaning Forgotten	Any other problems
MOAL01	N/A	Y	Y	75	3	Y	N	-
ABAL01	B1	Y	Y	100	3	N	N	-
SAAL01	A0 (High)	N	SOME	25	4	Y	Y	-
YUGE01	A0 (High)	Y	SOME	25-35	4	Y	Y	-
OMAL01	A1	-	-	25	3	Y	Y	-
ABAL02	C1	Y	Y	80-90	3	Y	Y	-
AHMO01	B1	-	Y	100	3	Y	N	-
RAAL01	B1	Y	Y	100	3	y	N	-
MOKH01	A2	Y	-	95	3	Y	Y	-

Speed of recording (speech rate) key:

1= I thought it was very slow, 2= I thought if was slightly slow, 3= It was just at the right speed,

4= I thought it was slightly fast, 5= I thought it was very fast

Speech rate per second: 1.88-2.12 words per sec. Avg.: 1.98

No. of words: 403

Length in minutes: 3.27

Understand general idea

All nine research participants were present for this listening lesson. Only one, SAAL01 said he did not understand the general idea.

Interesting

SAAL01 was in fact only one of two who said they found SOME of the passage interesting, whereas everyone else said they found the passage interesting.

%age Understood

SAAL01 indicated he understood only 25% of the passage. He had already stated that he did not understand the general idea, and found only some of the passage interesting. YUGE01, who also claimed to have understood 25-35% of the passage indicated that he understood the general idea. OMAL01 who also only understood 25% of the passage did not indicate whether he understood the general idea or found it interesting. The remaining participants understood from 75-100% of the message. Interestingly, ABAL01, AHMO01, RAAL01 and MOKH01 claimed to understand 95-100% of the listening passage, with CEFR grades of A2 and B1, yet, ABAL02 (CEFR rating C1) only claimed to have understood 80-90%.

Speed of recording

For SAAL01 and YUGE01, the speech rate was slightly fast, compared to everyone else who thought that the speech rate/speed was just right.

Any words not known and meanings forgotten

Although SAAL01, YUGE01 and OMAL01 claimed they only understood the passage in the range of 25-35%, they all indicated that they did not come across any new vocabulary. This seemed to at first contradict all of their other earlier responses. Yet, they also claimed that they came across words they already knew, but which they could not recall the meanings of. Five of the participants indicated that they came across familiar words but for which they couldn't recall the meanings, whereas four indicated no problems with words they already knew. So for this latter group, there was no problem in remembering or recognising or comprehending vocabulary already familiar to them. No other problems were indicated.

Table 21

*Class Profile 3, **MAIN IDEAS SECTION**, Book2, Unit 7, Listening 2-The Great Pacific Garbage Patch-BK2U7L2*

Research Subject	CEFR Listening Score	MAIN IDEA **TASK*** Difficulty	MAIN IDEA **TASK** Understood what to do?	MAIN IDEA **TASK** Understood Vocabulary	MAIN IDEA **TASK** All Questions Right?	MAIN IDEA **TASK** Any Other Problems
MOAL01	N/A	OKAY	-	SOMETIMES	N	ACCENT
ABAL01	B1	EASY	Y	Y	N	EASY
SAAL01	A0 (High)	DIFFICULT	N	-	N	-
YUGE01	A0 (High)	OKAY	-	SOMETIMES	N 2/4	KIND OF
OMAL01	A1	OKAY	SOMETIMES	SOMETIMES	N 2/4	**YES**
ABAL02	C1	EASY	Y	Y	Y 4/4	-
AHMO01	B1	EASY	Y	-	Y	N
RAAL01	B1	EASY	Y	Y	Y 4/4	N
MOKH01	A2	EASY	Y	Y	Y	ACCENT

***Task**: 3 option MCQ

Additions to Main ideas section

Additional questions that indicated the number of questions answered correctly in the main ideas section were added to the listening review sheet which provided further information about how learners got on in the classroom. It was felt that this would provide further useful information to the teacher. It is rarely possible to ascertain every learner's progress in class, however this would allow the teacher to gain further insight into how the learners were doing, and perhaps act as an additional starting point for further investigation.

Listen for main ideas task

Out of the nine, research participants, five thought the task was EASY, three thought it was OKAY, meaning it was doable, but it was not a simple or straightforward matter, and one person SAAL01 thought it was DIFFICULT and he did not know what to do. Five said they understood what they needed to do, with one (OMAL01) indicating he SOMETIMES knew what to do. This person also indicated that he only understood the vocabulary some of the times.

Of the five who thought the task was EASY, four got all of the questions right. Those who thought the task was OKAY, managed to get two out of four questions right. Bearing in mind that MCQ type questions are quite common in Saudi Arabia, it would seem that the main problem was simply not understanding what the instructions were, or probably more likely, what the answers or choices actually meant. Matching these up to a listening passage probably made it difficult. In the 'ANY OTHER PROBLEMS' column, OMAL01 mentioned that he did not understand any of the vocabulary. Despite this, he managed to get two responses right. Two research participants felt that the accent caused a problem.

Table 22

*Class Profile 3, **DETAILS SECTION**, Book 2, Unit 7, Listening 2-The Great Pacific Garbage Patch-BK2U7L2*

Research Subject	CEFR Listening Score	LISTEN FOR DETAILS **TASK** Difficulty	LISTEN FOR DETAILS **TASK** Understood what to do?	LISTEN FOR DETAILS **TASK** Understood Vocabulary	LISTEN FOR DETAILS **TASK** All Questions Right?	LISTEN FOR DETAILS **TASK** Any Other Problems
MOAL01	N/A	OKAY	Y	SOMETIMES	N	ACCENT
ABAL01	B1	-	Y	Y	N	-
SAAL01	A0 (High)	DIFFICULT	N	N	-	-
YUGE01	A0	OKAY	-	SOMETIMES	N 3/8	N

	(High)					
OMAL01	A1	DIFFICULT	SOMETIMES	N	N 5/8	Y
ABAL02	C1	OKAY	Y	Y	Y	N
AHMO01	B1	EASY	Y	-	N 7/8	N
RAAL01	B1	EASY	Y	Y	Y	N
MOKH01	A2	EASY	Y	Y	N 7/8	ACCENT

Task: 2 option sentence completion

Additions to Details section

As with the Main Ideas Section, additional questions were added to the listening review sheet relating to the number of correct answers.

Listen for details task

Three research participants found the task EASY, three found it OKAY, Two found it DIFFICULT. Six people had no problems with understanding what to do. One person said he sometimes knew what to do, with one person saying he did not understand what to do. SAAL01 and OMAL01 basically did not understand any of the task vocabulary, whereas MOAL01 and YUGE01 understood some of it. YUGE01 got three out of eight questions right, whereas OMAL01 despite claiming not to understand any of the vocabulary, managed to get five out of eight questions right. This indicated that there were other issues at play, or may be the answers to the questions were guessed. Two out of eight people got all answers correct. The issue of accent was brought up again.

Interviews

MOAL01

Whereas in Unit 7, listening 1, MOAL01 could understand 50% of the passage (See 4.3.1), he now said he understood 75%. He believed that words that were new in the first listening, he now knew, and thus these helped him to understand listening 2 better. The

question that arises, is whether those few words result in a 25% increase in understanding in the following passage. The estimation of what 75% may well vary with different learners, but calls into question this question as a measure of how much a learner really understands.

MOAL01 was not really able to explain why he had difficulty with the task other than a few vocabulary issues. He also mentioned 'accent' as an issue but could not recall why he mentioned this. I then played a short excerpt of the listening passage to see if this might result in a problem. On hearing the listening the only explanation he could come up with was that may be the sound was not good at the time.

ABAL01

Whilst interviewing ABAL01 about his responses, he explained that he did not know anything about this topic. He also said that he misunderstood the questions about 'ANY WORDS NOT KNOWN'. He had written down that he understood 100% of the passage, but initially put down 'Y' in this column, which indicated that he came across new vocabulary. After our discussion, he said that he meant NO, not YES. So then this was changed. This indicated that there was potential for misunderstanding the question on the listening review sheet.

With regards to the Main Idea and Details section, he stated that they were EASY, yet he did not get all of the questions correct, which he stated on the form. In trying to ascertain why he thought he was unable to get all of the responses correct, he explained that:
"I don't like to have many information in one class, my brain like goes away".
This could potentially either be a problem with remembering information, i.e. a short-term memory issue, but also relate to the spacing of task questions.

He also explained that it was possible that he did not listen to that part of the listening that was necessary to answer the question. I suggested that may be he was distracted and that may have been a cause of getting a response wrong, to which he agreed.

We then looked at the transcript for the listening passage, in which he had underlined what were new words, for example the word ‘shoreline’. I then asked again why he had said he understood 100% of the passage and said he knew all of the vocabulary if he had underlined a new word. He responded by saying that he did not hear it, and only saw it when looking at the transcript of the passage. Again, this seemed to bring up the issue of not hearing everything, and the issue of being distracted. He said quite clearly that this was a new word for him. In defending his position, he stated that:

“yeah I didn’t hear any new word, but when you gave me the sheet”.

In this sense, he answered the question on the listening review sheet correctly, as it asked if “did you hear any new words?”, which he clearly didn’t, but the transcript revealed that there were in fact some new words.

ABAL02

At the interview stage, ABAL02 reiterated his points about his mind sometimes wandering off resulting in missing out on some of the listening.

MOKH01

95% understanding meant that ‘not all the words’ were understood.

MOKH01 also felt that accent was an issue. What he meant (he said) was that some words were connected and that “did not pronounce ‘r’ and other letters.”

He also mentioned a phrase “on purpose” he came across. He knew the words ‘on’ and ‘purpose’ separately, but he did not know what ‘on purpose’ meant. This was a phrase that appeared in the listening passage.

SAAL01

I asked whether he thought the listening review sheet was useful. He said no. He felt that you *“just lose time”.*

I then tried to explain the purpose which included possibly helping him. May be holding on to the information for a month or so without the teacher taking action may have given that impression. He suggested that may be it benefitted higher level students, rather than him. SAAL01 seemed to think that the exam was *"better than the paper"*.

My interpretation was that he felt that it was more useful to do something that helped prepare for the exams. I tried to explain that this could potentially help to improve listening, and that without better listening, how was it possible to do well in the exam? SAAL01 said that he *"feel stupid"*. The process seemed negative.

Regarding the Main Idea and details sections the words were too difficult and he couldn't understand the pronunciation, or perhaps how to read the words. He then said he had problems with the grammar and vocabulary. Reading questions and instructions was a problem. However, he said that even if he could understand the instructions, he couldn't get the answer because he could not focus on the listening, one reason being that he had to keep an eye on the questions and also listen. He said that he understood the questions, though this seemed contradictory. Perhaps as the task required the selecting on a response from 2 options, he meant that he knew what he had to do. He basically felt that his English was too weak and this was the first time he was learning English in English.

A lot of time was spent on trying to convince him that this was a useful exercise. He had no comment, and preferred a more interactive approach. Nonetheless, he said he had a desire to speak English, but he was still not happy to learn English. The 25% he said suggested as a comprehension level actually meant 'very little'.

Having looked at class profiles based on the LRS, we now look at a sample of 3 individual class profiles.

4.3 Individual Profiles

As the listening lessons progressed, individual profiles gradually formed, with each lessons data being added to the previous lessons. It is these individual profiles that are now reported on. Although there were nine research participants, it was decided to select cases that represented a cross section of language proficiencies, as well those who had attended at least 8 of the 10 listening lessons. They also needed to have been interviewed.

The data for these research participants is presented in the same way as the classroom profiles, i.e. divided into three sections, with each section beginning with a personal profile. Additionally, responses for three listening passages are mapped to Field's Cognitive Framework.

4.3.1 Individual Profile 1 – SAAL01 (CEFR A0 High)

Aged 19, he was born in Madinah, Saudi Arabia and from an Arabic speaking family. His hearing was normal. He started studying English at school from the age of 11. He never attended any English classes outside of school and never spoke English at home. He did sometimes speak English with other people, but according to him, they would make fun of him. On a daily basis, he spoke English between 10-20% of the time, however, a large part of this was probably at the university, and during class time. In other words, it involved, reading, writing, speaking and listening. When asked about the amount of time spent listening to English via different forms of media, he claimed that this was 30-40% on a daily basis. He had never been overseas. In terms of accents, he had heard American British and Irish accents. These accents were encountered in the educational context from teachers. In terms of his views about the kinds of listening problems he had, he mentioned hearing words that he had perhaps heard before, but for which he had forgotten the meaning. He also mentioned issues of grammar, and not being able to understand words, possibly as a result of them being connected. He described his English level as okay. He felt that his weakest English skill was

speaking, followed by writing. Reading for him was the easiest skill, followed by listening. He took the Oxford Placement Test at the beginning of the research and scored CEFR A0, which is a sub-CEFR A1 score in the listening section. In the phoneme test, he scored 20 out of 28.

Table 23

Individual Profile 1 – Listener Perceptions - SAAL01 - CEFR - A0 (High)

Lesson	Understand General idea	Interesting	%age understood	Speed of Recording	Any words not known	Words Known Meaning Forgotten	Any other problems
BK2U6L1	Y	NS	35-45	4	Y	Y	-
BK2U6L2	-	N	0	3	Y	Y	-
BK2U7L1	-	-	-	-	-	-	-
BK2U7L2	N	SOME	25	4	Y	Y	N
BK2U8L1	Y	N	50	3	Y	Y	N
BK2U8L2	N	-	25	3	Y	Y	-
BK2U9L1	Y	-	75	3	Y	Y	Y
BK2U9L2	N	-	50	3	Y	Y	N
BK3U1L1	N	-	25	3	Y	Y	N
BK3U1L2	N	N	25	3	Y	Y	N

SAAL01 was one of the weakest students in the class and looking at the data above, especially the %age understood column, it seems that he struggled with comprehension for almost all of the passages but one. Out of the nine passages, he states that his comprehension was 50% or less in eight instances, with five being 25% or less. This information is useful and indicates that perhaps SAAL01 was in the wrong class, as things did not improve as the lessons progressed. Interestingly, for one particular passage, he claimed to have understood 75%. During the interview, he stated that he liked maths which was the topic of the listening passage (Appendix R), and that he was planning to do a maths degree in the following year. He stated that he liked maths a lot and that it was his favourite subject.

A question that does arise, is the meaning of 75. Just because of a topic that is of interest, SAAL01 gave it (BK2U9L1) a much higher rating than other passages. Looking at the lexical qualities of the passage (Appendix X), it can be seen that this passage in fact had the second lowest percentage of words from the Academic Word List (AWL). This is the same for the off-list %age of words too. There were also fewer words (Appendix Z) as a percentage with more than 2 syllables, compared to other passages, except one. So perhaps there was some justification for the higher %age understood figure indicated by SAAL01. However, the figure still seems much higher than one would expect. No statistic was available that referred to the percentage of words that were numbers within this listening passage. It was also on average one of the fastest passages (Appendix W), yet for SAAL01 the speed was just right.

For all of the passages, he stated that he came across new vocabulary and vocabulary that he knew but which he had forgotten the meaning of.

Individual Profile 1 - SAAL01 – Mapping to Field's Cognitive Framework

Table 24

Individual Profile 1: SAAL01 – CEFR A0 – Mapping to Field's Cognitive Framework

Perception Measures	Processes	BK2U7L2	BK2U8L1	BK3U1L2
Understand general idea	**Lower level Processes**			
	Input decoding	N	Y	N
	Lexical search	N	Y	N
	Parsing	N	Y	N
	Higher level processes			
	Meaning construction	N	Y	N
	Discourse representation	N	Y	N
Interest in Topic	**Lower level Processes**			
	Input decoding	SOME	NONE	N
	Lexical search	SOME	NONE	N
	Parsing	SOME	NONE	N
	Higher level processes			
	Meaning construction	SOME	NONE	N
	Discourse representation	SOME	NONE	N

%age understood	**Lower level Processes**			
	Input decoding	25	50	25
	Lexical search	25	50	25
	Parsing	25	50	25
	Higher level processes			
	Meaning construction	25	50	25
	Discourse representation	25	50	25
Speed of recording	Input decoding	4	3	3
	Lexical search	4	3	3
	Parsing	4	3	3
Words not known	**Lower level Processes**			
	Input decoding	Y	Y	Y
	Lexical search	Y	Y	Y
	Parsing	Y	Y	Y
	Higher level processes			
	Meaning construction	Y	Y	Y
	Discourse representation	Y	Y	Y
Words- meaning forgotten	**Lower level Processes**	Y	Y	Y
	Input decoding	Y	Y	Y
	Lexical search	Y	Y	Y
	Parsing			
	Higher level processes			
	Meaning construction	Y	Y	Y
	Discourse representation	Y	Y	Y
Other Problems: e.g. accent	Parsing	N	N	N

We now try to interpret the responses from the LRS in terms of Field's Cognitive Framework. All perception measures in the LRS are assumed to be the sum total of processing across lower and higher level processes. Therefore, in terms of diagnosis, these perception measures are at a surface level. It is not possible to break-down further based on the LRS data, where exactly in the listening process the problem is, and to what extent. Thus, for the first passage (BK2U7L2), SAAL01 indicates that he did not understand the general idea. This represents the sum total. Thus, at a lower level, there were probably issues at input decoding stage, as well as possibly a variety of lexical search issues, which could be as a result of a variety of factors such as word retrieval, segmentation, not recognising words. In terms of parsing, potentially there are a variety of issues, including issues of understanding

the structure, understanding the use of pitch and the impression one gets. Meaning construction would become difficult if there are problems at a lower level, and a representation of the discourse would be very limited. If we assume that all listening processes are at play simultaneously, then at even the surface level, the combination of problems with lexical searching, combined with the speed, and a lack of interest in the topic, results in a lower general understanding. For the second passage, despite a lack of interest in the topic, the claim is that the general idea was understood. Perhaps due to the speed of the passage being more manageable, this has helped in general understanding, though SAAL01 has stated that he understands more, suggesting greater success at the lexical search stage.

Table 25

Individual Profile 1 – Main Idea Tasks SAAL01 – CEFR A0(High)

Lesson	MAIN IDEA **TASK** Difficulty	MAIN IDEA **TASK** Understood what to do?	MAIN IDEA **TASK** Understood Vocabulary	MAIN IDEA **TASK** All Questions Right?	MAIN IDEA **TASK** Any Other Problems
BK2U6L1	HARD	-	N	N	-
BK2U6L2	HARD	-	-	-	Sound
BK2U7L1	-	-	-	-	-
BK2U7L2	DIFFICULT	N	-	N	-
BK2U8L1	OKAY	N	-	N	-
BK2U8L2	OKAY	-	N	Y	-
BK2U9L1	OKAY	Y	-	Y	-
BK2U9L2	DIFFICULT	N	-	N	-
BK3U1L1	OKAY	N	-	N	-
BK3U1L2	DIFFICULT	N	SOMETIMES	N	-

At no point did SAAL01 indicate that he found the main idea tasks easy, and almost for all passages indicated that he did not know what was expected of him, except in the case of the listening passage, in which he said he had comprehended the most, i.e. the passage related to maths. The task was a standard MCQ task, with three options. He did not seem to

indicate any other issues, except for a problem with sound, which he was not able to elaborate on at the interview stage.

Table 26

Individual Profile 1 – Detail Tasks - SAAL01 – CEFR A0(High)

Lesson	LISTEN FOR DETAILS - **TASK** Difficulty	LISTEN FOR DETAILS - **TASK** Understood what to do?	LISTEN FOR DETAILS - **TASK** Understood Vocabulary	LISTEN FOR DETAILS - **TASK** All Questions Right?	LISTEN FOR DETAILS - **TASK** Any Other Problems
BK2U6L1	HARD	N/A	N/A	N/A	N/A
BK2U6L2	-	-	-	-	-
BK2U7L1	-	-	-	-	-
BK2U7L2	DIFFICULT	N	N	-	-
BK2U8L1	DIFFICULT	N	N	N	-
BK2U8L2	OKAY	-	-	Y	-
BK2U9L1	OKAY	Y	Y	Y	-
BK2U9L2	DIFFICULT	N	-	N	-
BK3U1L1	OKAY	-	N	N	-
BK3U1L2	DIFFICULT	N	SOMETIME	N	-

Similar to the main idea tasks, SAAL01 seemed to have no problems with the one passage relating to maths, i.e. BK2U9L1. The task was a matching exercise, where statements that were listed on the right, had to be matched to numbers on the left. The numbers represented, distance, a phone number, a runner's number in a race, a house number from childhood, a number on a balloon, and time. SAAL01 felt very comfortable with this as already noted. Perhaps his recognition of numbers made it easier for him to answer these questions. However, in most cases, he seems to have struggled with the tasks in the details section. Overall, SAAL01 was a very weak student.

4.3.2 Individual Profile 2 – MOKH01 (CEFR A2)

Aged 18, he was born in Jouf, Saudi Arabia. He was also from an Arabic speaking family, where no one spoke English. His hearing was normal. He started to study English at school at the age of 13. He had not ever had any extra English lessons. He did speak English, but only during university class times. He said he spoke from 20-30% a day, which probably equated to the amount of time he was at the university. In terms of listening to English media, he stated that this did not exceed more than 10%. He had never travelled outside of Saudi Arabia. His experience of English accents, covered, American, British and Egyptian. These accents were experienced within an educational context. In terms of his perceptions of listening problems, he felt that issues included, accent, speed, new vocabulary, vocabulary known but forgotten, and issues with grammar, specifically part of speech. He also indicated that he might not recognise words in connected speech. He rated his English as bad, though he scored A2 on the OPT test. His weakest skill was writing according to him, followed by listening. Reading was the best skill, followed by speaking. He took the Oxford Placement Test at the beginning of the research and scored CEFR A2 in the listening section. In the phoneme test he scored 21 out of 28.

Table 27

Individual Profile 2 - Listener Perceptions - MOKH01 - CEFR – A2

Lesson	Understand General idea	Interesting	%age understood	Speed of Recording	Any words not known	Words Known Meaning Forgotten	Any other problems
BK2U6L1	Y	SOME	60-75	4	Y	Y	-
BK2U6L2	Y	Y	80-90	3	Y	Y	-
BK2U7L1	Y	Y	100	3	Y	Y	-
BK2U7L2	Y	-	95	3	Y	Y	-
BK2U8L1	Y	Y	85-95	3	Y	Y	-
BK2U8L2	Y	Y	75	3	Y	Y	-
BK2U9L1	Y	SOME	50	3	Y	Y	-
BK2U9L2	Y	Y	50	3	Y	Y	-

BK3U1L1	-	-	-	-	-	-	-
BK3U1L2	Y	Y	80-90	3	Y	Y	-

MOKH01 scored CEFR A2 on the OPT for listening. He understood the general idea for all of the topics, and found most topics interesting except for two. The first topic related to games that help children become more creative, and the second one related to numbers (Appendix R). Interestingly, compared to SAAL01, who scored much lower than MOKH01 in the CEFR, MOKH01 indicated that he only understood 50% of the passage (BK2U9L1) that related to maths. Making this comparison, it seems that perhaps the more interesting a topic is for the learner, the higher a %age understood figure is indicated and vice versa. To counteract this point though, MOKH01 indicated that he found the next passage interesting, yet he only managed to understand 50%. So there are probably other factors at play beyond just interest in a topic.

Taking a look at the data above, MOH01 seemed to require help primarily in the area of vocabulary recognition.

Individual Profile 2 – MOKH01 - Mapping to Field's Cognitive Framework

Table 28

Individual Profile 2 – MOKH01 - Mapping to Field's Cognitive Framework

Perception Measures	Processes	BK2U7L2	BK2U8L1	BK3U1L2
Understand general idea	**Lower level Processes**			
	Input decoding	Y	Y	Y
	Lexical search	Y	Y	Y
	Parsing	Y	Y	Y
	Higher level processes			
	Meaning construction	Y	Y	Y
	Discourse representation	Y	Y	Y
Interest in Topic	**Lower level Processes**			
	Input decoding	SOME	Y	Y
	Lexical search	SOME	Y	Y
	Parsing	SOME	Y	Y

	Higher level processes			
	Meaning construction	SOME	Y	Y
	Discourse representation	SOME	Y	Y
%age understood	**Lower level Processes**			
	Input decoding	70	90	85
	Lexical search	70	90	85
	Parsing	70	90	85
	Higher level processes			
	Meaning construction	70	90	85
	Discourse representation	70	90	85
Speed of recording	Input decoding	4	3	3
	Lexical search	4	3	3
	Parsing	4	3	3
Words not known	**Lower level Processes**			
	Input decoding	Y	Y	Y
	Lexical search	Y	Y	Y
	Parsing	Y	Y	Y
	Higher level processes			
	Meaning construction	Y	Y	Y
	Discourse representation	Y	Y	Y
Words- meaning forgotten	**Lower level Processes**			
	Input decoding	Y	Y	Y
	Lexical search	Y	Y	Y
	Parsing	Y	Y	Y
	Higher level processes			
	Meaning construction	Y	Y	Y
	Discourse representation	Y	Y	Y
Other Problems: e.g. accent	Parsing	N	N	N

Across the three passages, MOKH01 claims to understand the general idea. This is the output of the various processes across all listening perception measures, and represents meaning construction. To what extent the general idea is understood is difficult to ascertain precisely, but despite coming across new lexis, and also not remembering lexis already known, MOKH01 claims to have understood a great proportion of the passages concerned. He struggled with the speed of the recording of the first passage. To what extent this impacted on general understanding is not clear. Perhaps, he used pragmatic, contextual,

semantic and inferential information to help him. Overall, MOKH01 is a stronger listener than SAAL01.

Main Idea section (See Appendix L)

Out of the nine passages, MOKH01 found more than 50% of them easy. However, in one case where he indicates that the task is easy (BK2U7L1), he indicates problems with the task as well as the vocabulary. These responses seem contradictory. The task was a standard multiple-choice type, which Saudi's are well versed in, however, it is possible that the vocabulary made it difficult for him to answer the questions, which clearly seems to be the problem. Interestingly, he found the task to be easy for the last listening passage too, yet did not get all answers correct. So despite indicating that a task is easy, this does not necessarily lead to all task responses being correct, suggesting that there are other elements that need to be considered.

In the second passage, he indicates in the last column that perhaps the task responses were similar and that the amount of information made it difficult to respond to them. In fact, the task responses were in the form of 3 paragraphs, requiring learners to select the one closest to the main idea. It could be argued that to read a number of paragraphs in order to ascertain the main idea results in added cognitive load on many fronts.

He also states in two occurrences, that he found the accent challenging. A simple way of dealing with this may be to give him more exposure to the accent concerned, however, there could be other issues, for example pitch or rhythm. Further investigation would be required.

Details Section (See Appendix L)

MOKH01 found most of the detail tasks easy except for two. For the third listening passage, despite indicating the task was 'easy' he seemed to encounter other problems, and indicates problems related to vocabulary. For the fifth listening passage, he deems the task

merely ‘okay’, yet is able to get all responses right, despite his having issues with the accent. He finds the last passage okay, and again indicates problems with the vocabulary of the tasks, making it difficult to then respond as appropriate. These responses seem counterintuitive, and would need to be further investigated perhaps at interview time. Is it possible that the task looks simple, yet when the learner looks more closely at them, other problems, such as lack of understanding are involved? These points also highlight the possible shortcomings in the LRS which seemed to be able to capture certain data, yet there were apparent contradictions in learner responses, or perhaps other possible variables needed to be added so that more information could be captured. Importantly, perhaps what was being demonstrated here was the limitations of relying purely on student self-evaluation.

4.3.3 Individual Profile 3 – ABAL02 (CEFR C1)

Aged 18, and born in Madinah, Saudi Arabia, ABAL02 was from an Arabic speaking family. His hearing was normal. He started studying and learning English from the age of 12 and never attended additional English classes. English was never spoken at home. He spoke English only during class time with classmates and the teacher. On a daily basis, this equated to about 20-30% of his time. In terms of daily exposure to English via other sources such as social media, he selected 21-30%. He had never travelled abroad. In term of accents, he was familiar with American and British accents. These were in educational contexts, films, and soaps. Compared to the other learners, except for MOAL01, his exposure was wider, as the others had experienced the language mostly, or only in educational contexts. In terms of perceived problems with listening, ABAL02 selected accent, speed, and new vocabulary as problematic. He rated his English as good. He took the Oxford Placement Test at the beginning of the research and scored CEFR C1 in the listening section. In the phoneme test, he scored just 17 out of 28.

Table 29

Individual Profile 3 - Listener Perceptions - ABAL02 - CEFR – C1

Lesson	Understand General idea	Interesting	%age understood	Speed of Recording	Any words not known	Words Known Meaning Forgotten	Any other problems
BK2U6L1	Y	Y	85-95	3	Y	N	-
BK2U6L2	Y	Y	85-95	3	Y	Y	-
BK2U7L1	-	-	-	-	-	-	-
BK2U7L2	N	Y	80-90	3	Y	Y	-
BK2U8L1	Y	Y	90-100	3	Y	Y	-
BK2U8L2	Y	-	90-100	3	Y	N	-
BK2U9L1	Y	N	99-100	3	Y	N	-
BK2U9L2	Y	Y	100	3	N	N	-
BK3U1L1	Y	SOME	85-95	3	N	Y	-
BK3U1L2	-	-	-	-	-	-	-

Based on the OPT test, ABAL02 scored CEFR C1 (Appendix O), which made him the most proficient at listening in the class. A quick glance at the table above shows that he had no problems with listening passage speed issues at all. In that sense he was unique amongst the other research participants.

Looking at the table, he understood the general idea for all but one passage (which related to garbage), yet he found that passage interesting. He did not find one passage interesting (BK2U9L1), which interestingly was the passage about numbers which SAAL01 was very keen on. ABAL02 clearly stated during the interview that he did not like numbers, and that this was a weakness of his. He found the last passage he heard (BK3U1L1) somewhat interesting, which related to psychology (Appendix R).

In terms of %age understood, he was very reluctant to say that he understood a passage 100%. At interview stage, he made it clear that that there were times when his mind wondered off, or where he heard new words or came across vocabulary he had heard before but for which he could not remember the meanings. With regards to the one passage where he

indicated he understood 100%, he said that he fully focussed and his mind did not wonder off. Looking at the responses to this question by other learners who scored less on the OPT test, and thus had a lower CEFR score, they were nonetheless willing to claim that they understood 100% of a passage. This is consistent with other research (Luoma and Tarnanen, 2003; Jang et al., 2015) that demonstrates that lower level learners overestimate their ability, whereas higher level learners underestimate their abilities. ABAL02's responses also suggest that an element of caution is required when interpreting the data.

Individual Profile 3 – ABAL02 - Mapping to Field's Cognitive Framework

Table 30

Individual Profile 3: ABAL02 – CEFR C1 – Mapping a sample of Listening Perception Measures to Field's Cognitive Framework

Perception Measures	Processes	BK2U7L2	BK2U8L1	BK3U1L1
Understand general idea	**Lower level Processes**			
	Input decoding	N	Y	Y
	Lexical search	N	Y	Y
	Parsing	N	Y	Y
	Higher level processes			
	Meaning construction	N	Y	Y
	Discourse representation	N	Y	Y
Interest in Topic	**Lower level Processes**			
	Input decoding	Y	Y	SOME
	Lexical search	Y	Y	SOME
	Parsing	Y	Y	SOME
	Higher level processes			
	Meaning construction	Y	Y	SOME
	Discourse representation	Y	Y	SOME
%age understood	**Lower level Processes**			
	Input decoding	85	95	90
	Lexical search	85	95	90
	Parsing	85	95	90
	Higher level processes			
	Meaning construction	85	95	90
	Discourse representation	85	95	90
Speed of recording	Input decoding	3	3	3
	Lexical search	3	3	3
	Parsing	3	3	3

Words not known	**Lower level Processes**			
	Input decoding	Y	Y	N
	Lexical search	Y	Y	N
	Parsing	Y	Y	N
	Higher level processes			
	Meaning construction	Y	Y	N
	Discourse representation	Y	Y	N
Words- meaning forgotten	**Lower level Processes**			
	Input decoding	Y	Y	Y
	Lexical search	Y	Y	Y
	Parsing	Y	Y	Y
	Higher level processes			
	Meaning construction	Y	Y	Y
	Discourse representation	Y	Y	Y
Other Problems: e.g. accent	Parsing	-	N	N

Looking at the mapping above, it can be seen that ABAL02 did not understand the general idea of the first passage. This would suggest that although there was proficient listening at the decoding level, when it came to the higher level meaning-construction and therefore discourse representation, there were failures. Even though he states that the topic was interesting, higher level processing requires more information from within the listener. Did ABAL02 understand the intentions of the speaker? Was he able to apply personal general knowledge to the contents of the listening passage? If ABAL02 had difficulty with these issues, then perhaps he did not have the means that would allow him to select and integrate appropriate information. May be there was a memory issue, that needs to be considered. ABAL02 had already demonstrated his language proficiency in the OPT exam, and was the highest scoring of all the research participants. To ascertain why he had difficulty with the first passage would require further investigation.

Main Idea Section (See Appendix M)

ABAL02 did not seem to have any problems with the main idea tasks. This may suggest that in terms of problems, understanding the task and task requirements for somebody with a C1 on the CEFR is likely to be minimal.

Details Section (See Appendix M)

Although ABAL02 had no difficulty with the main idea task, he indicates that the task for passage 2 (BK2U6L2) was 'hard' and the task for passage 4 (BK2U7L2) was 'okay'. Taking a closer look at the tasks (Appendix R), listening passage BK2U6L, required the reading of 3 paragraphs before selecting the correct response. For the second passage, BK2U7L2, the task required learners to write notes based on prompts. Thus, this indicates that it is possible that there is a greater potential for more 'cognitive load' if there is no response to chose from. Looking at the class profile for this passage (4.2.2) , three other research participants who had a lower language proficiency, indicated that this task was either 'easy' or 'okay'. Interestingly again, a higher level learner still seems to underestimate his ability, or perhaps learner responses need to be clarified.

As stated already (see details section under 4.2.2) it was mentioned that ABAL02 wrote on the LRS that:

"They ask for numbers and I am not good in remember number and it was close the two number how say mention".

He also mentioned that:

"there were a lot of number, and I always forget number".

For him, *"it's about memory".*

Perhaps the questions about numbers overwhelmed him, despite there only being four out of nine questions that required information pertaining to numbers. Perhaps his memory was not that good, or perhaps the information required to complete the tasks were not sufficiently spaced out.

With regards to the listening passage BK2U7L2, during the interview stage, he felt that his mind wondered off, possibly affecting his ability to respond to the questions.

Looking at the lexical characteristics (Appendix X), this listening passage had the second highest percentage of words from the Academic Word list, thus potentially making the passage harder to listen to. At the same time though, this listening passage was also one of the slower ones in terms of average words per second (Appendix Q). Although ABAL02 was certainly one of the better learners, in the phoneme test that was taken prior to the research, he in fact scored below some of the much weaker students (see Table 3.1), which indicates that although he may not have been particularly proficient in listening to phonemes and differentiating between sounds, this did not necessarily it seems disadvantage him. Perhaps this needed to be investigated further. Did this may be indicate that he had issues with concentration?

Overall though, ABAL02 was a strong listener, based specifically on his percentage understanding figure.

4.4 Statistical analysis of listening review sheet data

The data relating to the listening review sheets is again presented in three parts as were the class and individual profiles:

1) Perceptions of characteristics of listening passage as indicated by student responses
2) Perceptions of the Main Idea section of the listening lesson
3) Perceptions of the Details Section of the listening lesson.

4.4.1 Descriptive and inferential statistics

Table 31

Descriptive statistics for LRS learner perception measures

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
InterestingTopic	58	3	1	4	3.17	0.13	0.994	0.987
Understood general idea	57	1	1	2	1.75	0.058	0.434	0.189
%age understood	75	100	0	100	64.787	3.7212	32.2265	1038.55
Speed	72	2	3	5	3.25	0.059	0.496	0.246
Words not known	75	1	1	2	1.827	0.044	0.3811	0.145
Meaning forgotten	75	1	1	2	1.627	0.0562	0.4869	0.237
Valid N (listwise)	47							

The table above provides overall descriptive statistics for this sample. As can be seen from the table, the mean for interesting topics was 3.17 which suggests that overall, the topics were close to interesting with a maximum of 4 on the ordinal scale.

In terms of the understanding of general ideas, the mean indicates that there was partial understanding of the ideas presented in the passages overall.

Percentage understood stands at a mean of 64.78 with a standard error of 3.72.

The speed mean indicates that on average, the speed of the passages were perceived at slightly above ‘just the right speed’. Thus, overall, the speed of a passage was found to be slightly fast for most of the research participants.

The words not known average indicates that more often than not, participants came across new vocabulary in virtually every passage.

Words heard and recognised as familiar, yet for which participants could not remember meaning came to approximately 1.63. This indicates that unless there is frequent exposure, words and their meanings are quickly forgotten.

We now examine each element in more detail, for each passage.

4.4.1.1 Interesting Topic (Appendix AB)

The passage with the highest mean is passage BK2U7L2 (n=7) with a mean of 3.71, which would indicate that this was the most interesting topic for participants. In contrast to this, the least interesting topic based on the mean, was BK3U1L2 (n=5) with a mean of 2.20, followed closely by BK2U6L2 (n=8) with a mean of 2.75. This information is important as topic interest is said to impact on comprehension ability (Buck, 2001; Bloomfield et al., 2010). Lack of topic interest may come about due to a lack of background knowledge and schemata, or even lack of vocabulary knowledge (Azmi et al., 2014), potentially impacting on working memory and thus in the understanding of the passage.

4.4.1.2 Understood General Idea (Appendix AC)

There were two responses available for this category, Yes (Y) or No (N), with Y=2 and N=1. Looking at the means in the table (Appendix AC), passages BK2U7L1 and BK2U9L1 both have maximum means of 2.00. This would indicate that all participants believed they understood the general ideas of the passages fully. Even SAAL01 who was one of the weakest said he fully understood BK2U9L1 which was about numbers. Interestingly, both of these passages involved conversations, with one being a radio interview (Appendix R). Looking at the remaining passages the next highest means are for passages where n=7, the mean being 1.86 for passages BK2U6L1 and BK2U8L1 with standard deviations of .378 respectively and variances of 0.143, indicating that the general idea of the passages was easier to understand than others. Interestingly, the standard deviations for passages BK2U6L2, BK2U9L2 and BK3U1L2 were .518, .577 and .548 indicative of a greater spread of the results for this category, suggesting may be that more participants found these passage difficult.

4.4.1.3 %age Understood (Appendix AD)

Looking at the percentage understood mean figures, it can be seen that the lowest mean relates to passage BK2U6L2 with a mean of 57.222, closely followed by BK2U9L2 with a mean of 57.857. These were as an average therefore perceived as the hardest passages. Taking the mean figures, it can be seen that passage BK2U8L1 was the easiest, with a mean of 74.375, followed closely by BK2U7L1 which had a mean of 72.143 percent understood.

4.4.1.4 Speed of Recording (Appendix AE)

In terms of speed, it can be seen that passage BK2U9L1 was considered by all participants to be just right in terms of speed, making it possibly the easiest to follow. However, this was not the slowest passage on average. In fact this was the third fastest passage based on average words per second (See Appendix Q). Seven other passages were slower on average and just two faster. The second ranked passage in terms of how learners perceived speed (meaning just right), was BK2U8L1 which was in fact the sixth fastest out of ten passages on average (See Appendix Q). BK2U9L2 the third ranked passage in terms of the speed that learners felt comfortable with , was in fact the slowest of all passages on average. These results seem confounding. However, BK2U9L1 and BK2U8L1, the first and second ranked in terms of speed just being right, both have the highest percentage of K1 words at just above 91% each (Appendix X) which could be a factor to consider in explaining this dichotomy between speed rate and learner perception. Passage BK2U9L1 also had greater content relating to numbers (Appendix Q). Did this play a part in the apparent ‘slowness’ of the speed that learners perceived? Even SAAL01, the weakest student seemed to think this passage was easy to understand. It is possible that his response impacted on the overall ranking of passages. This passage also had less than 6% of words with more than 2

syllables. Only one other passage BK2U7L1 had marginally less, but this was ranked as the third fastest by learners.

4.4.1.5 Words not known (new words) (Appendix AF)

In terms of new words, overall, the closer the mean is to 2, the more the indication that most found many new words in the listening texts. Passage BK2U6L2 was deemed by all to have new lexis. In fact this passage also has the largest percentage of Academic words from the Academic Word List (Appendix X) at 5.21% compared to other passages, though the type-token ratio was the lowest at 0.4 and lexical density 0.49 (Appendix X). BK2U7L1 was deemed to include less newer lexis by the learners, though in terms of academic words, it had 1.91%, and the most off-list words. (See Appendix X).

4.4.1.6 Words known but meanings forgotten (Appendix AG)

The higher the mean, indicates that participants came across many words that they had heard and come across in the past and may have been part of their lexical knowledge, but which at the time of listening, they could not recall the meaning of. Passage BK2U7L1 seems to have had words that everyone seems to have been familiar with and recalled without any problems. Thus there were no words that participants had difficulty with it seems. BK2U9L2 and BK3U1L1 were perceived to have a lot of vocabulary that was known in the past, but was forgotten.

4.4.1.7 Spearman's rho correlation (Appendix AH)

As can be seen from the table, significant correlations existed between 'understanding the general idea' and 'percentage understood', with a correlation coefficient of 0.675, being

significant at the 0.05 level. This was followed by a correlation of 0.488 between 'Interesting topic' and 'percentage understood.' In terms of inverse correlations, '%age understood' and 'meaning forgotten' had a direct inverse relationship and thus correlation of 0.561 which was significant at the 0.05 level. Thus, this seemed to indicate that these two variables have a correlation, where if more words were not known, the percentage understood was lower.

4.4.1.8 CEFR Correlations – Spearman's Rank correlation (Appendix AI)

The CEFR ratings correlated strongly with all of the measures on the LRS, indicating that the responses on the LRS reflected the language proficiency of the learners concerned, and thus showed a strong internal consistency in the LRS tool.

4.4.1.9 The Potential for diagnosis based the statistical analysis

Looking at the data based on the Spearman's rho correlation, it can be seem that there are potential relationships between the listening perception measures. If the topic is found to be interesting, then the chances are that a greater percentage of the passage is understood, which also suggests that learners are likely to understand the general idea of the passage. Where vocabulary previously known is forgotten, the percentage understood in a listening passage is also likely to be less. In terms of an initial basic diagnosis, the listening perception measures may provide some data for a teacher to understand why learners are having problems with certain listening passages.

Having illustrated and analysed class profiles, individual profiles and presented descriptive and inferential statistics, we now discuss the findings related to RQ1 (4.5.1) which is then followed by a discussion on RQ2 (4.5.2).

4.5 Discussion of findings related to research questions

4.5.1 Discussion of findings related to RQ1

RQ1 – To what extent is it feasible to diagnose listening problems in a classroom setting as part of every listening lesson within a Saudi Academic EFL context?

In total, there were ten listening lessons, with the LRS being distributed to all students during each class. This was done so that learners became used to the instrument so that they did not feel that this was anything extra. By distributing the LRS to everyone, the identities of the research participants were also not noticed, and everyone was seen as one body of students, doing the same things in class.

In a standard listening lesson, there are primarily three sections. Pre-listening, listening, and post-listening (Field, 2008). In this scenario, at pre-listening stage, listeners are encouraged to activate background knowledge possibly through questions asked by the teacher or in the book, followed by reading main idea questions, followed by playing a listening passage, and answering those questions. The listening passage may be played again so that learners can answer questions related to details questions. The teacher will then ask learners for their responses to the questions, and perhaps ask how and why they responded the way they did. However, the teacher has no further information to enlighten him or her about factors that may have resulted in difficulty for learners during this process. Thus, this more or less replicates a test process (Field, 2008).

The LRS was used in conjunction with the listening lesson as illustrated above. At the start, learners were asked to write down on the LRS in either English or Arabic what they thought the passage would be about and the possible words they might hear. The purpose was to activate schemata. If the learner was unable to write much, it indicated that perhaps the learner was not familiar with the topic. Although this data was available, this was not included in this research project. This was primarily because of time constraints. As the

responses required were in either Arabic or English, the amount of time required to analyse these would have stretched the research project in both amount of data produced and time required for analysis. This data will hopefully be analysed for a future project that takes a closer look at topic knowledge on it's own. After having gone through this 'pre-listening stage', listeners followed the same process as in any other listening lesson. Once the traditional post-listening stage was completed, i.e. ascertaining whether learners had managed to respond to the main idea and detail tasks, as a final part to the post-listening section, they were asked to fill in the LRS. The language instructor would then walk around to ensure that learners had no problems with filling in the LRS. In most classes, a transcript was also handed to the learners in order for them to ascertain if they had come across any new words, or words that perhaps they had known but forgotten, or perhaps even words that they had not heard. Once this was done, these were collected by the language instructor. The tool integrated well into the listening lesson.

Whereas in this research project the data has been presented in tables, a teacher would usually, just quickly look through the LRS sheets. As there were only 9 research participants, it was perhaps not such a difficult task. If there were let's say 30 to 35 students, then may be more time would have been required to look through the sheets. However, looking through learner responses would probably occur after the class, or while learners were engaged in another task on their own.

Depending on the responses, the teacher would then call certain individuals who it was deemed had difficulty with the particular listening passage, and then discuss with them their responses and any other problems they may have had that were not apparent from the LRS. In this project, the interviews took place after 5 weeks to facilitate the research project, however, in a more normal situation, learners would probably be seen in the same week, whilst the listening passage was fresh in their minds.

Compared to other research projects, there was no need for outside assistance. The teacher was able to review and collate the data, and interview individuals. There was no need for assistance during the lessons at any point.

In order to implement the tool, additional time was spent that would not normally occur had the LRS not been utilised in class. It could be argued that additional time could impinge on a normal listening lesson, however, the LRS seemed to integrate into the lesson and became a part of the normal procedure of a listening lesson. Overall, it was found that the LRS was well integrated into the listening lesson. Thus, implementation of the diagnostic tool in a classroom was feasible.

4.5.2 Discussion of findings related to RQ2

RQ2 – Do the resulting individual and class profiles raise an awareness of learners' strengths and weaknesses in EFL listening from a teacher perspective?

4.5.2.1 Class Profiles

From a teacher perspective, a glance across the class profile gave a quick overview of how learners had coped with the listening passages. The CEFR ratings for each learner acted as a kind of benchmark and a basic form of comparison between the learners. The CEFR ratings also resulted in an expectation of the kind of responses a teacher may expect from individual learners. Someone with a low CEFR rating should have more problems than someone with a higher CEFR rating. Thus one would anticipate language proficiency levels to be indicative of this form of expectation. This was the assumption when examining the class profiles. First one would look at all of each learners responses horizontally, and then down the columns, making swift comparisons in the process. Often, there was a tendency to look first at the %age understood column first, as this would provide immediate information on how learners seemed to have coped with each passage. Gradually, as more class profiles

were created, it became clear that there were particular areas where certain learners seemed to follow a particular pattern in their responses, and gradually as more class profiles were generated, it seemed that a certain set of learner responses seemed to fall within a certain range. This seemed to be an immediate benefit of looking at the class profiles. A swift glance provided an overview and a perceived impression of how the class had performed in the case of a particular passage.

In terms of understanding the general idea, there did not seem to be a clear pattern. However, those who had a language proficiency of A1 or less indicated at times that they did not understand the general idea. Nonetheless, this seemed to provide some evidence and scope for prediction from a teacher point of view, that certain learners may have a problem with the listening passage and thus potentially with other areas of the passage and even the tasks.

In terms of topic interest, there was no clear differentiator between the research participants. Nonetheless, it can be seen that there is a clear instance where despite a lack of interest, there was no difficulty in answering questions. ABAL02 who was the most proficient, had no issues with comprehension despite a lack of interest at times. This potentially confounds the view that lack of topic interest may impact on comprehension (Bloomfield, et al, 2010). Based on the data from the class profiles, perhaps this is something that applies more to lower level learners, rather than as language proficiency improves. In other words, lack of interest results in lower comprehension for lower level learners, but perhaps not for those with a much higher language proficiency.

Looking at the class profiles, more often than not, learners who had a listening proficiency of A2 and above, generally seemed to be able to understand more than 60% of the listening passages. Those below this generally struggled to understand even 50% of the passages. There were of course some anomalies. SAAL01 for example, claimed to have

understood 75% of a listening passage, justifying this by asserting that he liked the topic, i.e. mathematics. In terms of estimating one's own ability, it was found that one particular higher level learner; ABAL02, who had scored C1 on the CEFR, constantly underrated his listening performance. Others with a lower CEFR e.g. B1 would often claim to understand 100%, yet perhaps did not. Conversely, lower level learners at times overestimated their listening abilities. Nonetheless, this particular column stood out as providing a rough estimate of how learners coped with a particular listening passage, and thus helped the language instructor to predict how successful a learner was likely to respond to tasks correctly, especially if the learner had indicated that he understood only a low percentage of the listening passage. Again, learner perceptions of the listening passages were relied upon, and therefore so were any predictions of how successful task completion would be.

Speed of recording was also a very useful measure of learner perceptions towards passages, and again, there was a divide in terms of those who found the speed 'just right' as opposed to those who felt that the speech rate was 'slightly fast'. Again, those with A2 and above indicated less problems with speed. Thus, especially where lower level learners indicated that the speed of the passage was slightly fast, this also helped to predict success on the listening tasks. This factor would not normally be something that a language instructor would be aware of without a diagnostic tool.

With regards to speed or speech rate, what was also apparent, was that there were possibly other factors that affected speed perception. In one particular passage (See 4.2.2), it can be seen that two lower level learners who understood between 15-25% thought the speed was slightly fast, whereas a higher level learner who claimed to have understood 100% also thought it was slightly fast. If low level learners are not able to comprehend vocabulary, then perhaps this results in lower 'cognitive load' (Buck, 2001; Rost, 2011; Field, 2013). Had the learners been able to recognise some of the vocabulary, would this have resulted in higher

cognitive load? Would their response to the speed of the passage perhaps been different? From a language instructor perspective, this raised awareness of an issue that would not normally have been possible without a diagnostic tool. Thus, it became clear that viewing speech rate in isolation perhaps presented an incomplete picture.

Most learners found that they came across new words or lexis. Based on the class profiles, it can be seen that irrespective of language proficiency, everyone came across unfamiliar lexis. However, unfamiliar lexis may not be a problem, if the task does not rely on specific lexical knowledge. Perhaps as new listening texts are introduced, new lexis is inevitable. However, learners found this to be the case with almost all listening passages, suggesting that learners were being exposed to a large number of new lexis. This could be deemed normal, as learners were being introduced to new topics, with newer lexis. However, if the language instructor relies only on a language proficiency banding such as the CEFR or a test, in this case the OPT, then when the learner with the highest language proficiency is continuously indicating that he heard new lexis, this might call into question what the CEFR banding really reflects. From a diagnostic perspective, this is important as the tool helped to create an awareness that even higher proficiency learners have a lot to learn in terms of vocabulary and that it is the norm to come across new vocabulary. This had the effect of raising awareness of the limitations of apparently high language proficiency.

Many learners also indicating that they came across words or vocabulary, the meanings of which they had forgotten. This seemed to also be a recurring theme. Thus, looking at the class profiles, it could be seen that this was a ‘weakness’ that many suffered from, and for which some remedial exercise was necessary. Diagnostically speaking, perhaps listening passages which recycle the same or similar vocabulary again and again is what is needed to remedy this situation. Nonetheless, this created an awareness as a language

instructor, and looking at a class profile immediately brought this to the fore in many instances.

With regards to the main idea and details sections, the LRS questions relating to comprehension of tasks was useful. In many instances, lower level learners indicated that they did not understand what was required of them from the tasks. Many consistently indicated that they did not comprehend all of the lexis within the tasks. This would not be something that a teacher would be aware of in a listening lesson as a factor for not responding to a task correctly.

Even where tasks were understood in terms of concrete meaning, the need to then match these to a potential response from the within listening passage that had perhaps not been fully understood still made the task challenging. Thus, the overlap of understanding the task purpose, along with the lack of comprehension possibly resulted in incorrect responses.

The interviews demonstrated that learner's minds 'wandered off', and thus they did not hear every word. This perhaps reflects real life listening, e.g. when listening to airport announcements, where the traveller may focus on certain information, e.g. like the flight number, gate number, or time of arrival or departure, reducing the focus on co-text (Buck, 2001; Field, 2008). It could be deduced that listening is an activity where learners 'tune-in' and tune-out', purposefully, possibly in an unplanned way, though within the classroom, it is necessary to tune-in and tune-out to search specific information. This raises teacher awareness in the sense that perhaps understanding 100% of a passage, in other words, everything is not required. Perhaps most of the listening passage (i.e. not 100%) needs to be heard rather than all of it. In terms of the LRS, this is an important point to bear in mind, that not all of a listening passage needs to be heard, which perhaps is why even if ABAL02 (CEFR C1) was seen to underestimate his listening ability, though in reality he probably heard enough to respond to the tasks.

Some learners had also indicated problems with answering questions, indicating that there was insufficient space between what one task required, and what the next task required in terms of a response. Thus the information for a couple of tasks was perhaps 'bunched-up' all together, making it difficult for listeners.

Overall, the class profile was a useful way in which to obtain an overview of learner difficulties, and where in particular the problems or weaknesses seemed to be. Furthermore, as class profiles related to particular listening passages, they also provided an overview of learner perceptions about particular listening passages. Importantly, the information also allowed the teacher to ascertain which learners needed attention, and possibly a follow-up interview or meeting, in order to gain further insight into their responses, including perhaps looking into how the relevant and specific weaknesses could be addressed, which was the purpose of the class profile. This was thus a starting point for ascertaining which particular individuals were having problems with the listening passages. However, the compiling of more class profiles also resulted in the development of individual profiles, which shed more light on particular learner profiles which we now look at next.

4.5.2.2 Individual Profiles

Whereas class profiles provided an overview of how a group or class of learners perceived a particular listening passage, as well as giving a sense of the different levels of learners, the individual profiles enabled the possibility of looking at learners from an individual perspective.

Having individual profiles based on their responses to the LRS provided an overview of data about each research participant, something similar to a patient's record to which a doctor may have access.

SAAL01 (4.3.1) was possibly the weakest in the class, and it was clear that he had problems with most of the passages as indicated by the percentage understood column. As a result, he was not able to understand the general idea for most of the passages. This basic information clearly shows he was struggling, and thus it was relatively straightforward that the teacher could predict problems with the tasks, which would have more to do with just understanding the language of the tasks. Thus, as an initial marker or 'red flag' the percentage understood figure signalled that this learner had weaknesses in terms of passage comprehension.

MOKH01 who scored A2 on the OPT test had fewer problems, again the percentage understood column provided an indicator of how difficult the passages were. There were difficulties with some passages but overall, he was able to cope with most of the passages. He indicated a lack of interest in in one of the passages. A diagnostician would need to look at the passage concerned and ascertain what qualities resulted in the problems of comprehension, but in addition to this, an interview would clarify further the responses given by the learner.

Looking at the higher level learner (ABAL02), he also claimed to have problems. Perhaps the main issue was a lack of interest in some topics, and the other being the reluctance to say that he understood everything in a passage. What is clear from his profile, is that in terms of the data produced from the LRS, compared to other individual profiles, he is a strong listener. However, it was the interview stage that revealed more about why he had chosen the responses he had. Despite no apparent serious problems, interviewing him did bring to light issues that bring further light to the complexity of the listening skills.

The attempt to map the individual responses to Field's Cognitive Processing Framework for Listening was useful to a limited extent. Primarily, the listening perception measures need to be seen as the output that results from both lower and higher level

processes. Therefore, whereas the listening perception measures could be mapped to the framework, to ascertain at a deeper level the exact problems requires further investigation. Nonetheless, reconfiguring the individual profiles into a table that reflects Field's framework, gave the researcher time to reflect on what possible problems may lie beneath the problems indicated by learners.

Overall, the more data that is present, the more it is helpful to understand where individual weaknesses seem to generally lie. A bigger picture of the learners is also developed. By comparing the responses to a particular passage, with others before and after, it is possible for the language instructor or diagnostician to ascertain where an individual is having difficulties. Importantly, it was found that an interview stage is a necessity after having collected data, as further explanation by the learner introduces further insight to responses in the LRS.

4.5.2.3 Inferences from descriptive statistics of LRS (Listening Review Sheet)

The data analysis (Appendix AH) which focussed on the 6 variables (see 4.4.1) relating specifically to listener perceptions, provides useful information . It seems that topic interest did impact on understanding the general idea as well as percentage understood. The moderate correlation may be as a result of stronger learners not necessarily having a problem even where there is a lack of interest.

The numbers also confirm the close relationship between how much was understood in percentage terms and being able to understand the general idea, which was reflected by a correlation of 0.675, being significant at the 0.05 level. A question could be raised about the possibility of a halo effect between these two variables, however, a way to ascertain any such effect could be through discussion at interview stage. Meaning forgotten (with an inverse correlation of 0.351, significant at the 0.05 level), and new words (inverse correlation of

0.263, significant at the 0.01 level) also correlated significantly with 'understood the general idea'. This seems to be consistent with learner responses relating to vocabulary. Percentage understood correlated highly with interesting topic, understanding the general idea, speed, and new and forgotten lexis, with correlations being significant at the 0.05 level (See Appendix AH). Thus, again, there was consistency in the way learners were responding to the LRS.

Perception of speed on its own inversely correlated with percentage understood at 0.358 being significant at the 0.05 level. However, as has been shown above, percentage understood also interacted closely with the other variables, demonstrating that speed is just one element of many that impacts on comprehension.

Overall, these statistics seem to indicate that the use of figures from the LRS, make judgements about learner strengths and weaknesses plausible. However, what may impact on these numbers is the population being researched. A mixed ability group may result in a lower correlation. Focussing on particular learner levels may help to raise the correlation, positively or negatively.

In summary, and based on the points made relating to class profiles (4.5.1), individual profiles (4.5.2) and also taking into account the descriptive and inferential statistics (4.5.3), an awareness of learner strengths and weaknesses certainly does occur, providing the teacher valuable information about learners that would not otherwise be available without such a tool, and which could potentially be acted upon as a starting point for finding remedial solutions in response to this diagnostic data.

4.5.3. Discussion of findings related to RQ3

We now take a look at the possible link between the LRS and the lexical qualities of the listening passages.

RQ3 – To what extent are learner responses to the listening review sheet (LRS) related to the lexical characteristics of listening passages from an EFL academic text book?

This research question was added to the research project in order to provide additional data which may lend validity to the use of the LRS tool used in this project. Thus, in order to attempt to answer this question, a variety of lexical measures were needed, that would also somehow relate to the measures that had been used in the LRS. The six measures that were analysed and taken from the LRS were the following:

- Understood general idea (1=No, 2=Yes)
- Interesting Topic (1=Not sure, 2= No, 3=Some, 4=Yes)
- %age Understood
- Speed (1=Very slow, 2 = Slightly slow, 3=Just Right, 4=Slightly fast, 5=Very fast)
- New Words not known (1=All words known, 2=New words)
- Meanings Forgotten (1= NOT forgotten, 2 = Forgotten)

From the measures above, those that could realistically be compared to lexical measures were those which were conducive to measures of lexical difficulty. The category of 'interesting topic' could not really be compared to any measure. Topic interest is difficult to measure and as far as I know there is no measure which can measure the quality of a topic interest. With regards to words previously known and meanings forgotten, perhaps this is something that is related more to a learners own disposition. Thus this measure was also left out. Thus, it was felt that the following measures could perhaps be compared to a lexical measure.

- Understood general idea (1=No, 2=Yes)
- %age Understood
- Speed (1=Very slow, 2 = Slightly slow, 3=Just Right, 4=Slightly fast, 5=Very fast)
- New Words not known (1=All words known, 2=New words)

These variables were ranked based on LRS data, and then compared to lexical measures using the Spearman's rank correlation coefficient. The table below gives the figures relating to the correlation workings:

Table 32

Correlation of LRS Perception Measures with Listening passage lexical characteristics

		Ranking Understood Gen Idea	Rank Percent Understood	Rank New Words Not Known	Rank Speed
Spearmans rho	Correlation Coefficient	1.000	1.000	1.000	1.000
	Sig. (2-tailed)				
	N	10	10	10	10
K1RANK	Correlation Coefficient	-.320	-.030	-.013	.079
	Sig. (2-tailed)	.367	.934	.973	.828
	N	10	10	10	10
K2RANK	Correlation Coefficient	-.209	.042	**.557**	-.018
	Sig. (2-tailed)	.562	.907	.095	.960
	N	10	10	10	10
AWL	Correlation Coefficient	-.160	**-.576**	.156	**-.671***
RANK	Sig. (2-tailed)	.659	.082	.666	.034
	N	10	10	10	10
OFF LIST RANK	Correlation Coefficient	.012	.261	.206	**-.720***
	Sig. (2-tailed)	.973	.467	.567	.019
	N	10	10	10	10
TYPE	Correlation Coefficient	-.121	-.049	.239	**-.598**
TOKEN	Sig. (2-tailed)	.740	.894	.506	.068
RANK	N	10	10	10	10
MTLD	Correlation Coefficient	-.215	**-.491**	**.619**	**-.567**
RANK	Sig. (2-tailed)	.550	.150	.056	.087
	N	10	10	10	10
VOCD_RANK	Correlation Coefficient	-.382	**-.830****	.213	**-.476**
	Sig. (2-tailed)	.277	.003	.555	.165
	N	10	10	10	10
Word	Correlation Coefficient	.135	**.382**	.056	-.079
Concrete	Sig. (2-tailed)	.709	.276	.877	.828
Percent_RANK	N	10	10	10	10
2SyllPerc_RANK	Correlation Coefficient	**-.566**	**-.479**	-.219	**-.591**
	Sig. (2-tailed)	.088	.162	.544	.072
	N	10	10	10	10

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

4.5.3.1. Understood general Idea

There were no correlations that were significant at either 0.05 or 0.01 levels of significance. However, there was a moderate correlational relationship with listening passages that had a higher percentage of words with 2 or more syllables.

4.5.3.2. %age Understood

The figures show that the correlation is significant at the 0.01 level of significance with the Vocd lexical density measure, indicating that the higher the level of density, the harder listeners seems to find the passage. There were also moderate correlations with the AWL (Academic Word List) measure, indicating that passages with a higher proportion of academic word resulted in greater difficulty for the listeners. There was a moderate correlation with the MTLD lexical density measure at 0.491, and with passages having a greater percentage of words with 2 or more syllables.

4.5.3.3. (New) words not known

There were no correlations that were significant at either at the 0.05 or 0.01 level of significance. Nonetheless, there were moderate correlations with passages that had more K2 level words, as well as where the lexical density was greater based on the MTLD lexical measure.

4.5.3.4. Speed

The correlation is significant at the 0.01 level between speech rate and the AWL measure at 0.671, and between speech rate and off-list words at 0.72. There were moderate

correlations where there were more token types at 0.598, and the lexical measures , which is reflected also in the next lexical measure MTLD moderately correlating at 0.567, and a moderate correlation with the VOCD lexical measure at 0.476. Again, there is a moderate correlation with passages that have a greater percentage of words with 2 or more syllables.

4.6 Summary of Results chapter

This chapter has presented the results in order to answer the research questions in the following order: RQ2, RQ1 and RQ3. The results were first addressed in relation to research question 2. LRS data in the form of class profiles were first described from a teacher perspective, including the use of interview data to explain why learners had provided certain responses. Secondly, a sample of individual profiles were presented, with a description of the data at face value, along with interview data where applicable. Research question 1 was then addressed, describing the process of using the diagnostic tool and the practicality of implementing this in the classroom by a single teacher. Although this could have been presented first, it was felt that providing class and individual profiles, as well as interview data first, would demonstrate the amount of effort put into implementing the project. Addressing RQ1 after RQ2 would allow the readers to have an idea of what was being implemented. Lastly, RQ3 was addressed, taking into account lexical measures for the listening passages, looking at the possibility of any links between listening passage characteristics and learner responses based on the LRS. In the next chapter, we discuss the results, their implications and suggestions for further research.

Chapter 5 Discussion of results and conclusion

5.1 Synopsis

The purpose of this research was to investigate the feasibility of being able to implement a diagnostic assessment process within the listening classroom. An action research approach was necessary in order to understand the execution of such a process by a language instructor. This would be the foundation upon which the research project would be undertaken. Central to the diagnostic process was the listening review sheet (LRS), a research tool embedded into the classroom/listening lesson with the role of seamlessly fitting into a normal lesson, and with the ability to collect data that could possibly help the language instructor in identifying or diagnosing listening problems at both a whole class as well as individual level. Data from the LRS was analysed and converted into class profiles, individual profiles, and then described through the eyes of a language instructor. Learner interviews provided additional information to clarify learner responses, and were also used to provide further description to the class and individual profiles. The next step in the process was to use descriptive and inferential statistics to gain further cognizance into the LRS data and possibly discover relationships between certain responses and variables within the LRS. Lastly, a lexical analysis was undertaken of the listening passages in order to compare these with the learner responses in the LRS, with a view to discovering correlations that may lend validity to the use of the LRS as a diagnostic tool in the classroom. I now discuss the findings of the research.

5.2 Discussion related to findings

The overarching research question first:

5.2.1 RQ1 - To what extent is it feasible to diagnose listening problems in a classroom setting as part of every listening lesson within a Saudi Academic EFL context?

This question was important because traditionally in the language testing literature, diagnostic assessment, has been posited as something that is complex, time-consuming and difficult to implement, and very little research has been conducted into diagnostic listening assessment (Alderson, 2005).

The Listening Review Sheet (LRS), a tool developed by the researcher, was an attempt to realise the diagnostic tool described in principle two of the 'Tentative principles for diagnostic SFL assessment' (Alderson et al., 2014, p.20; Harding et al., 2015, p.318).

According to principle 2, the 'instrument' (2.6.3) should be designed to be user-friendly. The LRS was in the learners L1, as well as their L2. Thus, it gave learners the ability to respond to questions in either English or Arabic. Learners thus did not have to worry about not understanding what was required of them.

The instrument also needed to be 'targeted, discrete and efficient' (Alderson et al., 2014, p.20; Harding et al., 2015, p.318). The LRS set out to run parallel to the traditional listening lesson which comprised of, the pre-listening, listening and post-listening stages. In trying to explain the construct (see 2.6.2) that was adopted for the research, the listening lesson and its constituent parts were taken into account to ascertain which areas needed to be targeted. As Vandergrift (2007), stated, the product of listening in the classroom comes about as a result of not just the process of listening, but also as a result of non-listening factors, but which may not be clear to the teacher. Thus, in terms of what was reported in the research, the six variables that were classified under 'listener perspectives on the listening passage' (2.6.2) i.e.

- Interest in topic (Bloomfield et al., 2010)
- Understanding of the topic
- percentage understood
- Speech rate (Bloomfield et al., 2010)
- Lexis/vocabulary: (Field, 2008, p.87)
    - Words not known

    - Words previously known but forgotten
- Other: e.g. accent

were in alignment with the part of the listening lesson where the listening passage was played, and which asked learners to respond with their perceptions of the listening passage. As already stated, the variables targeted were based on the teacher's experience of the kinds of issues that learners might perhaps encounter in the classroom as well as issues that learners may easily be able to relate to.

With regards to the possibility of ascertaining problems that were non-listening in character, the next part: 'Listener difficulty with task prompts'(2.6.2), attempted to capture problems that learners would have with the tasks, which related primarily to whether learners comprehended tasks. Thus, the areas targeted were:

- Lack of understanding of vocabulary within task requirements (Field, 2008)
- Perceived difficulty of task

The LRS (Appendix B) tool attempted to gauge if learners had difficulties with the task aspect of the listening lesson as well. If research participant responses indicated problems with the task, this immediately gave the teacher extra information that would normally not be available. This raised the possibility that something beyond just listening process factors were causing problems for learners.

According to principle 2, what is targeted needs to be discrete. The question of what is or what is not discrete is an area of debate (Alderson, 2005, Lee, 2015), however, Lee talks about different levels or layers and the need to be open in how we define these, and Alderson talks about the purpose dictating what is discrete. As the purpose was to try and target elements of the listening lesson, that is the processes and tasks that are contained within the three stages in a typical listening lesson (pre-listening, listening and post-listening), it could be argued, that this was as discrete as one could be, as the target needed to have meaning in the eyes of the teacher, as well as the learner. The diagnosis that occurred was at the

beginning or surface level. In effect this replicated the initial diagnosis that occurs when for example a patient is visiting a doctor and where the doctor enquires of the patient about the kinds of problems they believe they have. The information produced acted more as an indicator that needed further investigation, especially as the data produced was based on learner's own perceptions of their problems. Thus, if one assumes that there are gradations of diagnostic data, then one could argue that within the context of a regular listening lesson, the data produced would provide leads or indicators to potential problems that require further investigation by a teacher/diagnostician. Where the learner states that he/she does not understand a passage due to a lack of familiarity with a topic, or where the listener lacks relevant vocabulary knowledge both of the listening passage and the task instructions, these provide concrete examples of problems encountered during the listening lesson process. The next step (not part of this research) would then be to ask the learner concerned to take further relevant diagnostic tests that are performance based rather than being only learner-perception based. Thus, the level of diagnosis that occurred here could be classed as an initial diagnosis, and from which a teacher or diagnostician would need to take further steps in order to confirm or clarify where problems actually exist. Not all problems may have the same level of difficulty to resolve. For example, dealing with task understanding may be a relatively easier problem to diagnose and fix, where specific vocabulary can be targeted for remedial work, whereas other issues relating to listening problems may require much further diagnosis.

According to principle 2, diagnosis must be efficient and suitable for administration in a classroom. The experience of using the LRS was that it integrated into the listening lesson and was designed to run concurrently with the traditional listening lesson plan. In terms of how seamless the integration was, the LRS was designed to cover all aspects of the three stages of the listening lesson, meaning the pre-listening, listening and post-listening. Thus, it was designed so that it complemented the lesson in the text. In an ideal world, it would be

included within the text book. Perhaps the non-integrating aspect was that the LRS was presented as an additional sheet of paper that learners had to fill in as they went through the stages of the listening lesson. Additionally, although a finite sample of learners participated in the research, every student was given an LRS and had to fill in the information required. This indicated that at least in a class of twenty five students, the LRS could be implemented. During the class, the teacher would walk around the class and glance at what students were writing on the sheets. However, it was once the LRS was collected from each student in the class that the teacher would take a more considered look at the information produced. This would happen outside of that particular lesson time. The actual analysis of the LRS data would occur after the lesson; in preparation for a teacher-student conference for those who the teacher felt really needed help. The teacher-student conference (or interview) would ideally occur within a day or two of the lesson. As already stated, the research context only allowed for interviews after all LRS data had been collected, which was after a period of approximately five weeks. Although the timing of the interviews during this particular research process was not ideal, the experience of having conducted the research would indicate that as long as teacher and student schedules allowed, interviewing and discussion of the diagnostic data would be feasible, thus allowing for further diagnosis and treatment.

The diagram below (figure 5) illustrates how the LRS diagnostic tool fitted into the listening lesson, based on a typical listening lesson. On the left hand side is the typical listening lesson with an additional stage at the bottom. The process of using the LRS tool is on the right. The arrows indicate when the LRS was used during the listening lesson. The bracket indicates that this section of the LRS was used during the whole stage, specifically what students remembered about what they heard, and which they were encouraged to write down.

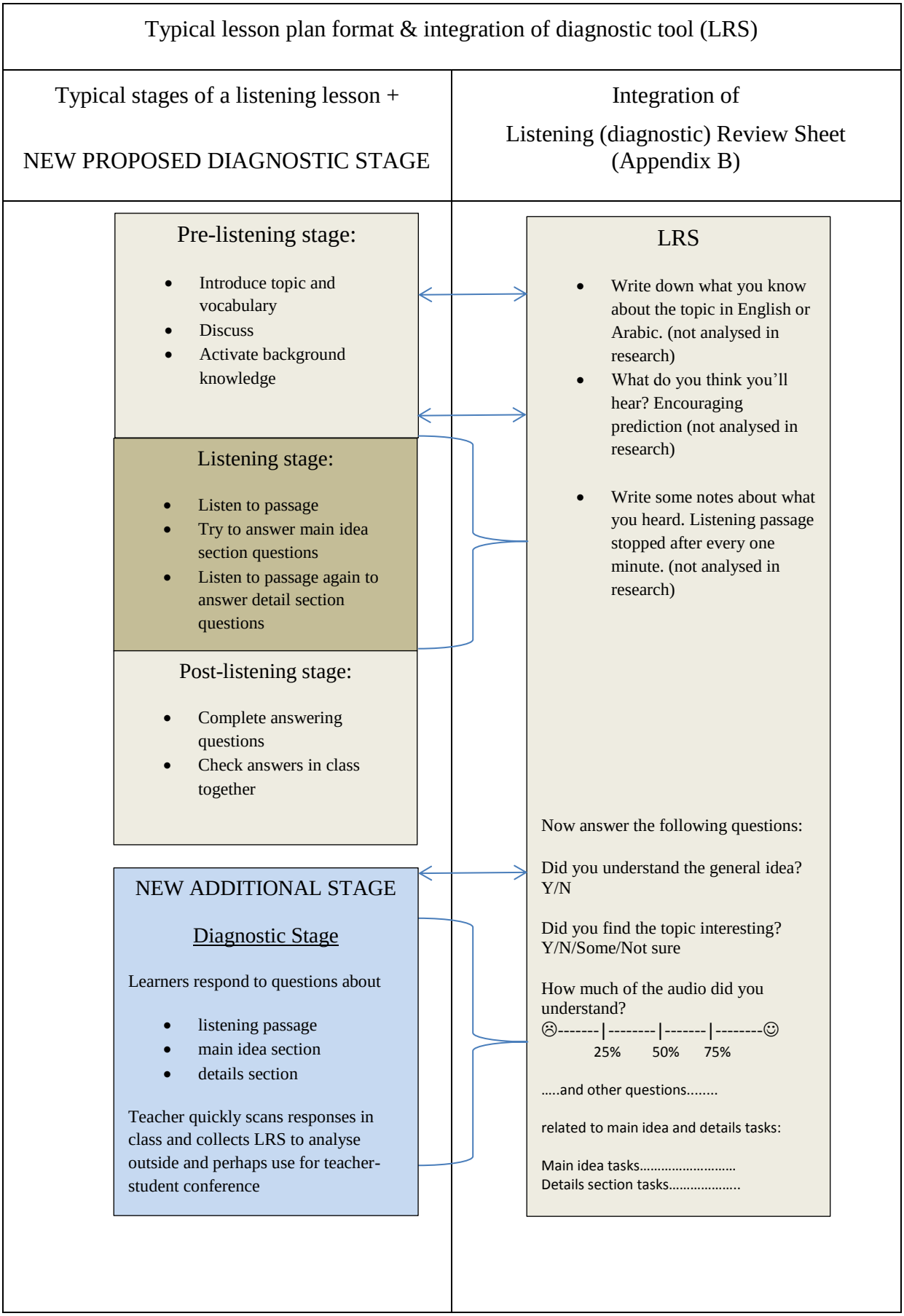


Figure 3 Typical lesson plan format & integration of diagnostic tool (LRS)

After the post-listening stage, in which the listening lesson task questions were completed and answers checked together in class, came the diagnostic stage (shaded in blue). At this stage, learners answered the 'diagnostic' questions in the LRS which related to the listening passage just heard, and the main idea and details tasks just completed. This additional stage provided the time and space to collect important additional potential diagnostic information (via the LRS) about listener perceptions that would not normally occur.

It was relatively simple to use, had a question format/self-assessment survey style, and relied on learner perceptions of the listening passage and the tasks they had to undertake. It was used for ten lessons and was designed so that it integrated into the standard listening lesson. There was no impact on the pacing schedule either, as ultimately, the job of the language instructor/researcher was to ensure that the curriculum pacing schedule of the institute was delivered on time.

The teacher was able to get an idea of learner perceptions of the listening passage by briefly looking at the LRS responses whilst walking around the class; however he later collected the LRS sheets. Thus, after the class ended, the teacher had enough information to allow him to examine the LRS sheets and ascertain how the class had done generally and individually, as indicated by the classroom and individual profiles. There was thus sufficient information to allow for the second stage of the diagnostic process, namely "*initial assessment"* (Harding et al., 2015, p. 318).

The profiles created as a result of the data from the LRS certainly provided a significant amount of data that gave the teacher an idea of where problems lay. However, the profiles were not shared with the "test-takers" (Harding et al., 2015, p. 323). In this context, there were no test-takers; they were simply learners in a classroom, using a diagnostic

instrument, which did not resemble a test in the conventional sense. The tool needed to be teacher generated, which it was, and thus the LRS fulfilled most aspects of principle 2.

As the tool was used from the start of the listening lesson cycle, learners were encouraged to indicate on it how familiar they were with the topic they were about to listen to, which in effect became part of the pre-listening stage. They were encouraged to write notes on the LRS when the listening passage was stopped intermittently, and then after having responded to the main idea and detail tasks, were asked to answer questions that related directly to the diagnostic information that has been presented in this research. Thus, the LRS was involved from the very start, and ended after the normal post-listening stage, integrated and used in parallel with the normal listening lesson. Analysis of the notes relating to the familiarity with the topic were not analysed due to time constraints, though there is potential to analyse these in a future research project. Rather, the listening perception question relating to topic interest was relied upon as this was deemed more time efficient for the purposes of this research. The predictive notes and notes written by the learners when the listening passage was stopped, could also potentially analysed, but were also not analysed due to time constraints.

In terms of breaking the normal listening lesson cycle, the LRS also opened up the possibility of teachers being able to understand why learners were perhaps not able to answer the task questions. As Harding et al. (2015) say "Of what practical use is it for a teacher to know that a student needs more help with 'listening for general ideas' (p.11). The LRS provided the potential for insight into why this might be the case. The multiple threads of information such as whether or not the learner found the topic interesting, the percentage of the listening passage learners perceived they understood, learner views on the speed of the listening passage, their perception of the words they comprehended or did not, and any other difficulties such as difficulties due to accent, all helped to provide additional insights. An

Issue other than pure listening but which nonetheless was part of the listening teaching process, was learner task reading problems for both main idea questions and detail questions. In other words, lack of familiarity with the vocabulary in the tasks also potentially impacted on the end goal of demonstrating comprehension and which could impact on actual task performance. These multiple areas thus provided potential lines of investigation for the teacher/diagnostician that would normally not be at the teachers disposal.

The explanation and information above is important to convey as the teaching of listening and assessment of it in the classroom has been something that has not changed much, and criticism of teaching listening as being merely test-like (Vandergrift, 2007;Field, 2008;Harding et al., 2015).

The LRS thus facilitated the possibility of revealing to an extent why some listeners were unable to respond to tasks correctly. Overall, taking into account that the teaching listening cycle still follows the 'test method' approach, this tool and its use in this research in my view is significant. Importantly, it provides an example of how such a tool could be implemented. Teachers, instructors or diagnosticians, could include appropriate targeted or discrete items to a diagnostic tool, and apply it any the listening classroom, depending on their purpose (Harding et al., 2015).

Whereas research does exist where learner diaries, verbal protocols (Goh, 2000) and other tools may have been used, to ascertain listener problems, the instrument used in this research was practical, and it is argued did not require any special time. As the researcher was teaching and researching at the same time, provision was made to ensure that allocated teaching time was not affected. All learners in the class used the LRS in class. This was necessary in order to ensure from a learner and classroom perspective, that the LRS was seen as a normal part of the lesson. Overall, there was no impact on the time required for completing the curriculum content prescribed and the material was covered in at least the

same depth. Thus, the data produced was ready to be turned into profiles that would give a teacher at the very least, a beginning idea of how the class in general, and individual learners were coping, and the more the LRS was used, the more information became available.

In terms of the diagnostic aims of the research there were two main aims. Firstly, a basic and feasible diagnostic assessment process that could be integrated into the existing format of teaching and assessing listening in the classroom. Secondly, such a process needed to be able to collect data that could aid the teacher in understanding learner strengths and weaknesses during the listening lesson. This data would raise teacher awareness of individual and class-wide learner problems, as well as provide a basis for further and deeper diagnosis, or even have the potential to help teachers develop a remedial plan. The LRS was able to fulfil these aims.

5.2.2 RQ2 – Do the resulting individual and class profiles raise an awareness of learners' strengths and weaknesses in EFL listening from a teacher perspective?

Whereas the implementation and practicality of a diagnostic instrument was the main focus of the RQ1, RQ2's focus was on providing valuable diagnostic information that had the potential to be acted upon, possibly for further investigation and follow-up with learners, and importantly in relation to a classroom context.

The type of information produced by the LRS provided an insight that would normally not be possible without such a tool. The results of the research and the subsequent class and individual profiles clearly show that it is possible to capture learner strengths and weaknesses. An important positive, is simply that what is not seen, could now be 'seen' (Vandergrift, 2007). This was however the beginning of the process.

Whereas the categories used in the LRS were specific, and for the purposes of the type of listening lesson being conducted, it required trial and error (Alderson, et al., 2014),

repetition of the diagnostic process, frequent looking back and forth of the data produced, and comparison, to understand what the information meant and what its implications were. This was a theme in the investigation by Alderson et al. into other professions in how they dealt with and approached diagnosis. As was stated, it was not just knowledge that counted; in fact it was experience too. Thus, initially, one would look at learner and overall class responses, and start making assumptions. When the next listening lesson resulted in data via the LRS, some assumptions about had gone before had to be revised. This was a natural tendency, as in terms of mind-set, a teacher (trying to learn to be, or act like a diagnostician) does not immediately think of a process model of listening when judging listeners. Perhaps this is a weakness in teacher training. The data that is presented at face value is the basis for any judgement.

A key indicator that acted as an anchor, was learner CEFR levels based on the Oxford Placement Test (OPT) that had been taken at the beginning of the research. These were present in all class profiles. These helped to provide an underlying assumption about learner listening proficiency. Perhaps if teachers have this information it might provide useful insight, especially if one has a multi-level class as was the case in this research. However, as listening is a complex skill, learner responses did not necessarily always reflect their CEFR proficiency level. Nonetheless, if someone had a CEFR level of C1, as an 'experienced' teacher, I could use my 'intuition' (Alderson, et al., 2014), and experience of teaching higher level students to try to understand why a particular response was given. Similarly, if a lower level student, for example someone who had an A0 or A1 on the CEFR, the chances were that there were other or different issues at stake.

For example ABAL02, who was a C1, consistently under-rated his performance especially in terms of how much he understood of a listening passage. Learners who had lower CEFR levels, e.g. B1 or B2 were consistent in rating their understanding of the

listening passage as higher than ABAL02. SAAL01 proved this, by rating his understanding of a listening passage about maths significantly higher than any of the other passages, just because he liked maths and numbers. Thus, this reflects the research on self-assessment (Unaldi, 2014), where there are certain tendencies to overate one's own language ability especially at lower level, and possibly under-estimate ability at a higher level. This was something that became prominent in my mind as a teacher/diagnostician as the diagnostic process progressed, as I had gained some experience of looking at the data. If anything, this made me slightly more cautious in how I approached judgment of the data and learners specifically. Again, the process of knowledge vs experience was slowly inculcating its way into my mode of analysis of the data, which is seems consistent with the accounts of people from other professions. In attempting to tap into particular areas (Alderson, 2007) to use as a basis for diagnosis, these were useful.

Knowing that the learner may not be familiar with a topic, or may lack interest in a topic, immediately acted as a warning sign; and as has already been discussed in the literature, (Sadighi & Zare, 2006;Tyler 2001), less familiarity with a topic does result in less comprehension. So as a teacher learning to use a diagnostic tool, using teaching experience and knowing the learner, based on their CEFR level, or previous LRS responses, based on previous listening passages, there would be an expectation about whether the learner would understand the topic, and how successful they were likely to be with the e.g. the main idea and detail tasks. If a lower level learner indicated a lack of interest in a topic, then one would expect that this person was probably going to encounter difficulty; however, if a higher level learner indicated a lack of interest in a topic, then there would be a different, probably more positive expectation.

So in the case of the weaker student, who expresses a lack of interest in a topic, one would assumes that there was going to be (using Field's (2013)) cognitive framework for

example), less input decoding, and as has already been suggested, this would mean difficulty during a lexical search, as well as problems with parsing. So failure at a lower level, would probably mean failure in the listeners mind to construct any meanings, and success in creating a bigger picture. Trying to decipher important points from the less important points (discourse representation), would probably never happen. In addition to this, one would expect problems with speed too. On the other hand, a higher level learner who indicates a lack of interest in a topic, may not necessarily have these problems. These were nuances that were noticed as more data was collected, and assisted in developing thinking about how different listeners listen, and which possibly relate to their language proficiency.

In terms of intuition, the percentage understood figure was usually the first 'measure' looked at. This also set expectations of how listeners were likely to cope with the listening passage. As someone who has taught listening for many years, there was information available to hand, that provided further insight into the learners in the class, and which resulted in forming opinions about general learner traits as more data was collected.

Although there was a desire to understand both strengths and weaknesses, there was a tendency to seek out weaknesses. From an instructor point of view, if a learner is seen to be relatively strong in a particular area, there may be a tendency not to think about the strength any further, as the focus may be on just finding the problem, and 'fixing' it. Lee (2015), is of the view that the purpose of diagnosis is to find weaknesses, and then to find remedies for them. The LRS was created to ascertain strengths and weaknesses, and was designed to try and reveal both. However, an the underlying assumption was that weaknesses were to be the focus of attention, and not the strengths. Ultimately, whether one should focus on either the strengths or the weaknesses, depends on the users of such tools, but the LRS provided data that acted as an indicator for both.

5.2.3 RQ3 – To what extent are learner responses to the listening review sheet (LRS) related to the lexical characteristics of listening passages from an EFL academic text book?

The correlational links with the LRS data were strongest between speed of recording (speech rate) and the lexical density measures. In this sense, learner perceptions do seem to have had some validity in this sense, and I would suggest concur with e.g. Nation (2001), who makes claims about the link between lexical knowledge and listening fluency. This does not in any sense mean that one could say that the LRS is totally, reliable; however, it could be argued that it is a start, and others could potentially develop the idea further, by tapping into other targeted and 'discrete' measures that suit their purpose and context.

5.3 Significance of the research

This research was novel and original in a variety of ways.

5.3.1 Addressing the diagnostic framework principles

The Harding et al. (2015) framework (Please see Table 1) was a useful tool and allowed the researcher to plan how diagnosis could potentially be realised in a classroom context. Of importance, was Principle 4 which comprises four stages which reflect the steps a teacher/diagnostician may take in order to ascertain learner strengths and weaknesses. These are discussed below.

1. Listening/observing stage. According to Harding et al. (2015), a teacher may be able to develop a deep or 'detailed' (p.13) understanding of an individual learner's listening problems through observations in the classroom. This may include interacting with the student informally too. However, this researcher's experience of teaching is that not much can be ascertained merely through informal interaction. Furthermore, with typically a large

number of students in the classroom, or even a few, ascertaining individual listening performance is quite challenging.

As is suggested by them (Harding et al., 2015), to make a more formal diagnosis, the teacher can provide a checklist which allows learners to self-assess their strengths and weaknesses. Learners can also be provided an appropriate task that assesses their listening comprehension, which may be followed by a meeting outside of class to ascertain specific difficulties.

The LRS required learners to respond to self-assessment questions about the listening text. In terms of self-assessment questions relating to the actual listening, this occurred immediately after administering a listening task that focused both on main ideas and specific details. Thus, students perceptions of their listening experiences were more or less real-time, and during the listening lesson. As the LRS was used over a 5 week period starting from the beginning of the semester, it was used as if it was a normal part of the listening lesson.

As was suggested by Harding et al. (2015), a teacher-student meeting was also held. Thus, using the responses to the LRS, the teacher was in a position to ascertain and seek clarification on the kinds of listener problems learners believed they had. Thus, the teacher/diagnostician was able to develop some hypotheses based on student responses, and the subsequent discussion. As the data collected over a five week period resulted in class and individual profiles, this provided the teacher/diagnostician with several or more examples of perceived learner difficulties. The LRS questions also fitted into Field's listening framework (See Table 8).

2. Initial assessment. Based on the above, a teacher/diagnostician should be able to hypothesise about the kinds of difficulties that learners seem to be having (Harding at al., 2015).

The focus of the LRS was both on listening, as well as on learner ability to cope with the tasks used to respond to main ideas and specific details. In terms of listening, the teacher/diagnostician was able to ascertain at an initial level, that learners were having problems with (for an example see 4.3.1 - SAAL01) lexis (for example see Table 23: 'the percentage understood',' words not known' and 'words known meanings forgotten'), and speed of speech (speed of recording). These broadly fell into Field's listening framework, with speed impacting on input decoding, and problems with vocabulary aligning with lexical search problems, both of which potentially impact on parsing. Not being able to understand the general idea as well as learner interest in the topic could also be located in Field's framework, however at a much more general level. Percentage understood indicated that underlying problems including lexis and speed impacted on parsing, which warranted further more precise investigation.

3. Hypothesis checking. Harding et al. (2015) suggest that at this stage, a teacher/diagnostician should have access to tasks that confirm any hypothesis formed. No further tasks were available to the researcher in this context, and thus no further tasks were given in order to verify the initial hypothesis.

However, the researcher suggests that the class and individual profiles which resulted from the data collection cycles that occurred over five weeks, act as a form of validation. The repeated collection of data to an extent confirmed the difficulties that learners were facing. Nonetheless, no further tasks which would result in independent performance data were used during the research.

4. Decisions making. According to Harding et al. (2015) the teacher/diagnostician would need to decide whether the initial hypothesize was correct and supported by evidence. As already stated in the previous point, the classroom and individual profiles provided some evidence, as the data collected related to 10 cycles of data collection over a period of five

weeks. This helped the teacher/diagnostician ascertain if there was any consistency in the responses of the learners. For example, if a learner stated that there were problems with the speed of the recording, the individual profile was able to provide a consistent record of this. Furthermore, learner responses relating to problems with lexis were relatively consistent. These provided substantial data indicating a problem. For a teacher, this data would provide sufficient evidence that learners were having problems with the speed of texts, especially those who had a lower CEFR level. In this sense, the teacher was confident that the learner responses pointed towards a particular problem. Consistency in response through the data collection cycles was confirmation of a hypothesis, at least in the mind of the teacher in a classroom context.

Feedback and a follow-up are also required at this stage. In essence, this was not really provided. At the interview stage, what resulted was clarification, but no real feedback other than perhaps agreement between the teacher/diagnostician and learner, that there existed problems relating to lexis and speed.

Of the four points discussed above, the first two were clearly addressed in the research, with the third according to the researcher addressing to an extent hypothesis checking due to the availability of data over a ten cycle process. With regards to point four, no follow-up procedure was put in place due to the limitations of the research aims.

The tentative diagnostic framework set out by Alderson et al. (2014), as far I believe has not been tested out in a real context as far as is known, and thus in this sense, this research tried to add to the vision that the authors had in mind. Although the framework is tentative, in nature, it has the potential to be changed or added to.

The experience of having designed a diagnostic tool based around the diagnostic principles was useful in the sense that the framework acted as an initial guide. However, if it is to be used in a classroom context, more flexibility will need to be added. For example,

which stage comes first, second or third should be left open to allow for experimentation. Often, with a class of thirty students, it is not possible to know for sure which students are struggling and which are not. In this research, it was the LRS that basically provided more detail which potentially alerted the instructor to problems that even those who seemed fine were having. The terminology also should be changed. Terms like testing do not really fit into a classroom environment. Assessment is more appropriate for this environment, as it alludes to wanting to help learners improve. These are however, minor details. The framework acts as a good starting-point for further research and experimentation. To help it become more well-known, actual examples of it's use need to be available to present to teachers in a variety of contexts.

5.3.2 Introducing diagnosis as the norm in the listening lesson

As Alderson (2005) stated in his book, *"*Only through the trial and error of developing diagnostic instruments, based on both theory and experience of foreign language learning, are we likely to make progress in understanding how to diagnose, and what to diagnose" (p.25).

I believe that the listening review sheet (LRS) used in this research was successful to some extent in meeting Alderson's view, and I believe that its successful implementation can be realised in the listening classroom. Thus, implementing the LRS in the listening lesson also resulted in what I believe, to be an important break from the normal way in which listening is taught in the classroom. The LRS as a tool fitted into the normal: pre-listening, listening and post-listening phases. It was possible to integrate it into the lesson, and be able to collect information about the learner which would not normally be possible. It focussed on areas that learners could easily relate to, and which seemed to have direct relevance to what they could perceive.

Although the listening lessons I taught during the research followed very much the lamented 'testing model' (Field, 2008), the LRS was able to perhaps take advantage of this process and produce information that would create an awareness both for the learner, but especially the teacher. Thus, the testing model was modified in the sense that learner information about the difficulties they were having was made available. Perhaps listening teachers in different contexts, not just that teach an academic listening lesson, or even one in the Middle East, could try out the LRS. As the three stage listening lesson is something that is followed in most listening lesson contexts, the LRS was designed to fit into this. What is required, is for teachers to try out the idea and then perhaps for them to share their experiences. This can then be fed back into the general teaching of listening.

Thus, diagnosis does not have to be in the form of an explicit test, rather it can just be a tool that is not seen as a test, and is accepted without any form of stress. Also, with the current norm of non-intervention, language instructors may have the opportunity to become interventionists in a class setting that is not used to this role being taken on by the listening teacher. In this sense, I believe the research has perhaps introduced an idea that could be considered for future implementation within the listening lesson cycle.

5.3.3 Characteristics of listening

Some of the listening characteristics that were used for the LRS resulted in some interesting observations. With regards to passage topic, although there is a view that topic familiarity, prior exposure (pre-listening in the classroom), academic, non-academic characteristics can impact on comprehension (Bloomfield et al., 2010), what was observed during this research project, was that language proficiency also impacts on the comprehension of topics. It was observed, and the data indicated, that a higher level learner who had no interest in a topic, or had come across something new, was still able to cope with

or deal with the passage at hand. This seems to align well with normal listening experience. Thus, this showed that even a lack of background knowledge or interest in a topic can still result in full comprehension; however, this level of listening would be a characteristic of a higher level learner.

In terms of speech rate, it was found that different learners had different perceptions of the listening passages, which of course linked through to their language proficiency. It seems that other factors play a role in speech (recording speed). In one instance, a student suggested that the speech rate was fine; however, he could not comprehend the passage. Yet, if he was in a position to try to comprehend words even at a lower level, this would probably have increased his cognitive load, and possibly resulted in a change of mind of how fast the speed of the recording was. This seemed to suggest that speech rate and information density may impact on listening comprehension, although this was not explicitly investigated here. The percentage understood measure, although qualitative and self-produced was a very strong indicator of listening comprehension, having a correlation of 0.774 (Appendix AI). I do not believe that this measure has been used before to measure listening passage difficulty from a learner point of view.

5.3.4 Field's Cognitive Framework for Listening

In an attempt to validate the individual learner profiles, Field's framework was used to map the listening perception measures. In reality, the perception measures within the LRS represented the end-point of a complex listening process. It was not possible with the data available, to ascertain beyond these broad labels. The underlying problems remained hidden, and require other forms of tests for them to be teased out. However, placing the perception measures within the context of a process model, even within a table, acted as reminder that there are many processes that exist within listening that are still not visible.

5.3.5 Action Research in the classroom

The desire to try out and introduce part of a diagnostic process into a daily classroom lesson and make it relatively straight forward to use by a teacher was an underlying aim of the research. The theory was already in place via the Harding et al. (2015) diagnostic framework. The LRS acted as the data collection tool in order to try and bring into practice some of these ideas. Just like a teacher may give learners a small quiz on a daily basis and review the quizzes later to ascertain how learners are doing, the LRS collected data of a different kind. It was this that formed the basis for creating individual as well as classroom profiles, which it was hoped would help the teacher/diagnostician ascertain possible learner problems.

To recap the classroom process (see also 3.8.3.1 and figure 3 under 5.2.1), the LRS tried to encompass the three listening stages of pre, during and post-listening and also added an additional stage to allow for data collection. This way, it would be seen as part and parcel of the listening lesson. Not just the research participants, but all learners were handed a copy of the LRS. All of the learners, and importantly research participants, would fill in the relevant information depending on the stage of listening. Perhaps the most important part in terms of data collection for this research was the additional stage that came after the post-listening stage. At this stage, listeners would fill in information relating specifically to their perceptions of not just the listening audio, but also any challenges they had in relation to actually comprehending the tasks. Thus an attempt was made to collect diagnostic data in real-time to ascertain potential listening problems and non-listening problems.
Walking around the classroom and glancing at the LRS in front of learners certainly allowed the researcher to get an idea of how learners had done. A quick glance at the percentage understood figure would provide an immediate, but rough impression of how particular

learners had found the listening passage. A quick peak at learners responses related to the tasks also provided an immediate idea of how successful learners probably were doing. A question could be asked of a particular learner of how they felt they had done whilst walking around. This provided information not normally available to a teacher whilst teaching listening. In contrast to this, if students are involved in writing something in the classroom, the teacher can merely through a glance, glean an idea of how the learner is doing because it is a productive skill.

After collecting the LRS in class, and whilst learners were given another task, the responses from the research participants were quickly skimmed and scanned, and an impression formed. Later, the data would then be entered into a spreadsheet. First, the class data was collated, and then individual profiles created. The class profile immediately gave an overall impression of how learners had coped with the listening passage that day. The responsibility of collecting the data, and then collating it, and reviewing it occurred within a short time span. Thus, the teacher was involved in all stages of data collection which also included reviewing and then analysis whilst and after the profiles were created. This was only really possible using action research. The teacher/diagnostician drew meaning from the data, being fully aware of what the learners had encountered in the classroom. In effect, this represented a cycle of events that were repeated ten times over five weeks.

As the lessons continued (and these cycles repeated), individual profiles grew in size. These provided the teacher/diagnostician more details, with patterns emerging of how specific learners were coping. This allowed the teacher/diagnostician to form and confirm views about the kinds of problems specific learners were encountering. Importantly, as more data was collected, this provided a confirmation or otherwise of what had gone before.

The interaction of the teacher/diagnostician with the data resulted in reflection and a greater awareness of what was occurring in the classroom. Not only did he deliver the lesson,

he was also a researcher. By understanding the theoretical background of listening, and diagnostic assessment, he was able to understand to what extent the theory could be brought into practice.

Action Research allowed the researcher to try and emulate a real teaching/learning situation in combination with an element of diagnosis. It requires more deliberation, organisation and reflection than normal by the teacher (Kemmis & Taggart, p.10, 1992). These ten cycles of organising, trialling, discovering and contemplation resulted in a greater awareness and understanding of learner problems within the limits of the LRS. Importantly, it not only created a greater awareness of learner problems every time a new cycle of data collection occurred, it also provided opportunities to reflect on how useful the data produced was, and whether it had real meaning.

Perhaps as a researcher, the biggest concern is whether implementing the LRS via Action Research is seen as a valid attempt at researching diagnostic assessment in the classroom. For example, without a third party involved, how does one deal with the issue of data validation? The position of the researcher here is that repeating the data collection cycle ten times provided a level of data validation. The scanning of the data as the profiles grew larger provided the teacher/researcher with information that allowed him to discover and recognise patterns of learner problems, resulting in greater awareness. The triangulation of the LRS data which was collected ten times, coupled with learner interviews provided further confirmation of learner responses. Lastly, aspects of the data collection tool linked through to Field's Cognitive Validity Framework (2013).

This research encompassed, diagnostic assessment, listening assessment and assessment in the classroom by a teacher. Action Research provided the opportunity to combine all of these. The complexity of the classroom as a research context however, is not

simple and far from perfect especially when trying to research a perception skill such as listening. Action Research made this possible.

5.4 Limitations of the research

As the LRS was in effect a self-assessment tool, and no performance data from an external test was used to verify learner claims, this could be deemed an important limitation of the research. It could be argued that no external checking took place in order to verify learner claims, despite the interview during which the researcher sought to verify learner responses on the LRS. The research relied very much on learner responses, and this data formed the basis for the results of this research.

Based on the feedback of SAAL01, it also seems that there is the potential for learners not to respond appropriately. His reluctance to fill in some information in Arabic on the LRS because of his embarrassment, despite his lack of English proficiency, may suggest that users need proper training before using such tools to ensure data accuracy. In fact SAAL01 was quite negative (see 4.2.3) about the LRS. If the LRS or a similar tool is to be used in any future research or within a teaching context, then the clear purpose of the tool and the buy-in of users is necessary. SAAL01 clearly did not see the benefits of the tool, and felt embarrassed using it. Perhaps this was because the research itself did not involve a remedial planning stage that could be based on the apparent difficulties he had. If a remedial stage had been added, this may have added to the perceived usefulness of such a tool for him. To ensure that the LRS is used properly, learners need to be clear about the purpose of the tool, training provided and any concerns need to be resolved. SAAL01's responses also seem to reflect the importance of an interview stage. Although it was primarily to discuss learner responses and delve more into problems encountered during the listening lesson, it was also useful in finding out about other learner problems that could impact on LRS responses.

The LRS only partly produced data that could be categorised under Field's Cognitive Processing model. There were other items that did not fit into this model because they were not of a listening nature. In this sense, for future research, perhaps a model is required that takes into account non-listening factors too.

The research did not set out to use performance data, as the research aims were limited to looking at the feasibility of a tool that would provide initial diagnostic data. Appropriate tests that would be able to test learner claims could be an important next step in the research process, however, not for this research. Having said this, it would be entirely appropriate to design tests that are diagnostic in nature and could test learner claims. Perhaps these could form the basis of additional research.

In an ideal situation, once a listening passage had been played in the class, and the LRS's collected from the learners, it would have been appropriate to have arranged a teacher-student conference soon after. Instead, in an effort to maintain research participant interest, I did not wish to burden them, by requesting them to take time from the heavy schedules. The loss as a result of this was that when five weeks later they were interviewed, many could not recall exactly what was in their minds when they filled in the LRS. Their memories could be jogged, but the time lapse may have inevitably impacted on more insights from the learners. These potential insights could have included their actual perceptions of the relevant listening passages at that particular point in time, as well as their ability to distinguish between the issues they may have encountered when dealing with various tasks. Thus, any future research would ideally include a teacher-student conference soon after the actual listening lesson in order to gain as much insight as possible into the learner experience.

The listening passages were not analysed in terms of their phonological characteristics. Rather, measures that one would normally use for measuring text qualities were used. These have been used in other listening research too; however, phonological

information about the listening passages could have added more information and possibly provided insight into learner perceptions. Tools such as PRAAT could be used, but due to a lack of knowledge on the part of the researcher, this tool was not used. There is generally a lack of tools that measure the characteristics of listening passages.

5.5 Implications.

The primary impetus of this research project was to respond to calls for further research that could help in the development of diagnostic assessment (Alderson, 2005; Harding, et al., 2015), an area that still requires much work to be done.

Very much related to the above, was a goal to see how assessment can play a part in the learning of a language, and thus its role in the classroom. This is an ideal that especially language testers have been discussing (Rea-Dickins, 2008; Turner & Purpura, 2015) for a while, but which has not been realised as perhaps the message still needs to reach the language teaching profession. Language testers in this sense are very much at a distance from language teachers.

As Alderson (2005) stated in his seminal work, most language testing research centres around large-scale testing. This research has added to the very few other examples of diagnostic assessment in the classroom (Doe, 2011; Fox & Hartwick, 2011), but has demonstrated I believe that a single teacher can create and implement a practical diagnostic tool, and use it in the classroom.

As diagnostic testing in this research was used to provide data that could help to alleviate learner problems, and ultimately improve their language learning and development, it should be seen as an additional tool that can be used in conjunction with assessment for learning (Black, Harrison, Lee, Marshall, & William, 2003), learning oriented assessment

(Turner & Purpura, 2015), and classroom-based language assessment generally (Rea-Dickins, 2008).

In order to be able to realise ambitions of bringing diagnostic assessment into the classroom, teacher training is necessary, and language testing trainers need to be able to understand the complexity of the language classroom context.

Materials writers and publishers should consider introducing a form of self-assessment within the texts that can facilitate the diagnosis of listening problems. Many texts now provide can-do statements based on the CEFR, adding a self-assessment tool that relates to other listening or indeed task related problems would make it easier for the language instructor to ascertain learner problems that would not normally be possible.

In addition to a transcript at the end of an academic listening text, a basic set of information about a listening passage would aid the teacher in predicting potential problems. Perhaps information relating to length of passage, speech rate, the type of listening passage, e.g. monologue or dialogue, academic or non-academic.

The provision of this information could help teachers become more aware of these characteristics and thus aid them in their development and awareness of listening passages.

Lastly, there needs to be a more pro-active effort from the language testing community to convey and transfer good testing and assessment knowledge, in order for these ideas (e.g. diagnostic testing) so that learners may benefit.

5.6 Future Research

This research did not consider the kind of remedial treatment or planning a teacher or language instructor would consider or implement after having received diagnostic information in relation to a learner. Whereas this researcher is a language tester and language instructor, most are not. In that sense, research that encompasses the development of tools

which could become standardised and used in particular contexts would certainly move the diagnostic assessment field forward. Additionally, tools that are easily implementable in low resource contexts would also be useful. Not every institute can have access to a language or computer lab. Therefore, tools that can be easily used by a language instructor in the classroom, and materials that can be used outside of the classroom, but which are easily accessible by both learner and teacher or diagnostician would help.

Most language instructors are practical people who need a practical tool that can be implemented, and are busy teaching listening, speaking, writing, reading, grammar and vocabulary to a vast array of students of all ages. Diagnostic tools and remedial programmes for such vast and diverse areas of knowledge such as these have yet to be developed, and disseminated across language teaching and assessment. There is potential for much research that can involve language instructors.

Language instructors need to be encouraged to take on action research projects that will allow them to research their classroom contexts when trying out or implementing diagnostic tools. This is necessary as it is not always easy for a third person to come in and do the research as an observer. In that sense, if there are people who teach, then they need to be encouraged to take-up research with the premise that the result could improve language learning.

5.7 Conclusion

I took on diagnostic assessment as a research project because of my position as primarily a language instructor who knows a reasonable amount about language testing, and who feels that assessment as a subject is not really well understood in the language teaching domain. Undertaking the research in the classroom was a purposeful decision as it is easier to relate the experience to language teachers who may see language testing as something that is

distant and perhaps even esoteric compared to their own experience of language teaching. Being seen as ‘one of them’ may open up opportunities to convey the message of language testing in a way that is nuanced and more relevant to teachers and their classrooms. Encouraging, funding and awarding scholarships for assessment research in the classroom, or rather teaching and learning (as testing and assessment may be seen as anti-learning) is the way forward (I believe) for language testing if it is to be genuinely influential and heard in the language learning field. Language Testing research needs to go beyond just researching high-takes testing (Alderson, 2006).

Bibliography

Alderson, J.C. (2005). *Diagnosing Foreign Language Proficiency*. New York: Continuum.

Alderson, J.C.(2007). The Challenge of (Diagnostic) Testing: Do we know what we are Measuring? In J.Fox, M.Wesche, D. Bayliss, L.Cheng, C.E.Turner & C.Doe (Eds), *Language Testing Reconsidered* (pp. 21-40). Ottawa: Ottawa University Press.

Alderson, J.C. (2010). Cognitive Diagnosis and Q:Matrices in Language Assessment: A Commentary. *Language Assessment Quarterly, 7(1),* 96-103. doi: 10.1080/15434300903426748

Alderson, J.C., Brunfaut, T. and Harding, L. (2014). Towards a theory of diagnosis in second and foreign language assessment: insights from professional practice across diverse fields. *Applied Linguistics*, 36(2), 1-26.

Alderson, J.C., Clapham, C. & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.

Anderson, J.R. (2009). *Cognitive Psychology and its implications* (7th ed.) New York: Worth.

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38*(4), 561-613.

Azmi, B.M., Celik, B., Yidliz, N., & Tugrul, M.C. (2014). Listening Difficulties Encountered by Students in Second Language Learning Class. *Journal of Language and Instructional Studies in the World, 4*(4), 1-6.

Bachman, L.F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L.F. (2004). *Statistical Analysis for Language Assessment.* Cambridge: Cambridge University Press.

Bachman, L.F., & Palmer, A.S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning*. Berkshire: Open University Press.

Bloomfield, A., Wayland, Sarah, S.C., Rhoades, E., Blodgen, A., Linck, J., & Ross, S. (2010). *What makes listening difficult? Factors affecting second language listening comprehension*. University of Maryland Center for Advanced study of Language. https://www.casl.umd.edu/wp-content/uploads/2016/02/WHAT-MAKES-LISTENING-DIFFICULT-tech-details-2011.pdf

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer* (version 6.0.25) [computer software]. Retrieved September 2016 from http://www.praat.org

Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing, 19*(4), 369-394.

Brown, G. (1986). Investigating listening comprehension in context. *Applied Linguistics*, 7(3), 284-302.

Brown, H. D. (2004). *Language Assessment, Principles and Classroom Practices*. New York: Longman.

Brown, J. D. (2005). *Testing in Language Programs*. New York: McGraw-Hill.

Brunfaut, T., & Révész, A. (2015). The Role of Task and Listener Characteristics in Second Language Listening. *Tesol Quarterly, 49(1)*, 141-168.doi:10.1002/tesq.168

Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing,* 15(2), 119-157.

Chang, A.C., & Read, J. (2006). The effects of listening support on the listening performance of EFL learners, *TESOL Quarterly*, 40(2), 375-420.

Chang, A.C-S, Wu, B, W-P., & Pang, J,C-L.(2013). Second Language Listening Difficulties Perceived By Low-Level Learners. *Perceptual & Motor Skills: Learning & Memory*, 116(2), 415-434.

Chapelle, C.A.,Cotos, E., C. & Jooyoung,L.(2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing, 32(3),*385-405.doi:10.1177/0265532214565386

Cheung, H. (1994). *The perceived and actual effects of speech rates on listening comprehension*. Unpublished master's thesis, City University of Hong Kong.

Churchland, P. (1999). Learning and conceptual change: the view from the neurons. In A.Clark and P. Millican (Eds), Connectionism, concepts, and folk psychology (pp.7-43). Oxford, UK: Oxford University Press.

Cohen, C., Manion, L., & Morrison, K. (2013). *Research Methods in Education*. London: Routledge

Cresswell, J.W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. California: Sage Publications Inc.

Cross, J., & L. Vandergrift. 2015. 'Guidelines for designing and conducting L2 listening studies'. *ELT Journal* 69/I:86-9

Coxhead, A. (2000). A New Academic Word List. *Tesol Quarterly*, 34(2), 213-238.

Council of Europe. (2001a). *Common European Framework of References for Languages: Learning, teaching, assessment*, Cambridge: Cambridge University Press.

Cumming, A.(2015).Design in four diagnostic language assessments. *Language Testing, 32(3)*,407-416.doi:10.1177/0265532214559115

Dahl, D. (1981). The role of experience in speech modifications for second language learners. *Minnesota Papers in Linguistics and Philosophy of Language, 7*, 78-93.

Davidson, F. (2010). Why is Cognitive Diagnosis Necessary? Reaction. *Language Assessment Quarterly, 7 (1)*,104-107. doi: 10.1080/154343300903426755

Derwing, T., & Munro, M.J.(2001). What Speaking Rates do Non-native Listeners Prefer? *Applied Linguistics*, 22(3), 324-337.

Doe, C. (2011). The Integration of Diagnostic Assessment into Classroom Instruction. In Tsagari, D. and Csepes I. (Eds), *Classroom-Based Language Assessment* (pp. 63-76). Peter Lang: Frankfurt.

Doe,C. (2014). Test Review - Diagnostic English Language Needs Assessment (DELNA). Language Testing, 31(4), 537-543.doi:10.1177/0265532214538225

Doe, C. (2015).Student Interpretations of Diagnostic Feedback. *Language Assessment Quarterly,12(1)*,110-135.doi:10.1080/15434303.2014.1002925

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

Elder, C., & Randow, J.V. (2008). Exploring the Utility of a Web-Based English Language Screening Tool. L*anguage Assessment Quartely.5:3*, 173-194, DOI: 10.1080.154343300802229334

Ferris, D., & Tagg, T. (1996). Academic listening/speaking tasks for ESL students: Problems, suggestions, and implications. *TESOL Quarterly,* 30(2), 297-317.

Field, J. (2008). *Listening in the Language Classroom*. Cambridge: Cambridge University Press.

Field, J. (2013). Cognitive Validity. In A. Geranpayeh and L.Taylor (Eds.) *Studies in Language Testing 35, Examining Listening, Research and practice in assessing second language listening*. (pp. 77-151). Cambridge: Cambridge University Press.

Flowerdew, J., & Miller, L. (1992). Student perceptions, problems and strategies in second language lecture comprehension. *RELC Journal,* 23(2), 60-80.

Fox, J., & Hartwick, P. (2011). Taking a Diagnostic Turn: Reinventing the Portfolio in EAP Classrooms. In Tsagari, D. and Csepes I. (Eds), *Classroom-Based Language Assessment* (pp. 47-62). Peter Lang: Frankfurt.

Fulcher, G. (2012). Assessment Literacy for the Language Classroom. *Language Assessment Quarterly,* 9(2), 113-132.

Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book*. London and New York: Routledge.

Graham, S., Santos, D., & Francis-Brophy, E. (2014). Teacher beliefs about listening in a foreign language. *Teaching and Teacher Education,* 40, 44-60.

Graham, S. (2017). Research into practice: Listening strategies in an instructed classroom. *Language Teaching,* 50(1), 107-119.

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System,* 28, 55-75.

Graesser, A., McNamara, D.S., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioural Research Methods, Instruments and Computers*, 36, 193-202.

Graesser, A.C., McNamara, D.S., & Kulikowich, J.M. (2011). Coh-Metrix:Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, 40(5), 223-224. DOI: 10.3102/0013189X11413260

Griffiths, R. (1990). Speech rate and NNS comprehension: A preliminary study in time-benefit analysis. *Language Learning*, 40(3), 311-336.

Griffiths, R. (1992). Speech rate and listening comprehension: Further evidence of the relationship. *TESOL Quarterly*, 26(2), 385-390.

Hasselgreen, A., Carlsen, C., & Helness, H. (2004). European survey of language testing and assessment needs. Part one: General findings. [Online]. Available at http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf (Accessed 27 November 2014).

Harding, L., Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing, 32(3), 317-336.* doi: 10.1177/0265532214564505

Harsch, C. (2014). General Language Proficiency Revisited: Current and Future Issues. Language Assessment Quarterly, 11(2), 152-169. doi: 10.1080/15434303.2014.902059

Hirai, A., & Koizumi, R. (2013).Validation of Empirically Derived Rating Scales for a Story Retelling Speaking Test. *Language Assessment Quarterly, 10(4),*398-422.doi:10.1080/15434303.2013.824973

Hirsch, D., & Nation, I.S.P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8,* 689-696.

Huhta, A. (2014). *Research Note: Diagnosing the development of writing ability*. Retrieved from http://pearsonpte.com/wp-content/uploads/2014/07/Hutha_A_2014.pdf

Isaacs, T., Trofimovich, P., Yu, G. & Chereau, B.M.(2015). IELTS Research Reports Online Series, *Examining the Linguistic aspects of speech that most efficiently discriminate between upper levels of the revised IELTS Pronunciation scale*. ISSN 2001-2982, Reference: 2015/4. Downloaded on 2nd September 2015.

Jang, E.E. (2009a). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for a Fusion Model application to LanguEdge assessment. *Language Testing, 26 (1),* 31-73.

Jang, E.E. (2009b). Demystifying a Q-Matrix for Making Diagnostic Inferences about L2 Reading Skills. *Language Assessment Quarterly, 6(3),* 210-238. doi: 10.1080/15434300903071817

Jang, E.E. (2010). Demystifying a Q-Matrix for Making Diagnostic Inferences about L2 Reading Skills: The Author Responds. *Language Assessment Quarterly, 7(1),* 116-117. doi: 10.1080/15434300903559225

Jang, E.E., Dunlop, M., Park, G., & van deer Boom, E.H.(2015).How do young students with different profiles of reading skill mastery, perceived ability, and goal orientation respond to holistic diagnostic feedback? *Language Testing,32(3),*359-383. doi:10.1177/0265532215570924

Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing, 27(4),* 555-584.

Johnson, B., & Christensen, L. (2012). *Educational Research, Quantitative, Qualitative, and Mixed Approaches*. Thousand Oaks, California: Sage Publications Inc.

Kang, O. (2013). Linguistic analysis of speaking features distinguishing general English exams at CEFR levels. *Cambridge English: Research Notes*, 52, 40-48.

Kemmis, S. and Taggart, R. (1992). *The Action Research Planner (Third Edition)*. Geelong, Victoria: Deakin University Press.

Kemmis, S. (1993). Action Research. In M. Hammersley (Eds). *Educational Research: Current Issues* (pp. 177-190). London: Paul Chapman Publishing.

Kim, A-Y. (2015). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing, 32(2),* 227-228.doi:10.1177/0265532214558457

Kim, Y-H. (2011). Diagnosing EAP writing ability using Reduced Reparameterized Unified Model. *Language Testing, 28(4),*509-541.doi: 10.1177/0265532211400860

Klimova, B.F., & Hubackova, S. (2013). World Conference on Psychology and Sociology 2012, Diagnosing Student's Language Knowledge and Skills. *Procedia – Social and Behavioral Sciences,* 82(2013), 436-439. doi.10.1016/j.sbspro.2013.06.289

Knoch, U. (2009).Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing, 26 (2),* 275-304.

Knoch, U. (2012). At the intersection of language assessment and academic advising:Communicating results of a large-scale diagnostic academic English Writing assessment to students and other stakeholders. *Papers in Language Testing and Assessment, 1*, 31-49.

Koizumu, R., Hideki, S., Ido, T., Ota, H., Hayama, M., Sato, M., & Nemoto, A. (2011). Development and Validation of a Diagnostic Grammar Test for Japanese Learners of English. *Language Assessment Quarterly, 8(1),* 53-72.doi:10.1080/15434303.2010.536868

Lee, Y-W.(2015).Future of diagnostic language assessment. Language Testing, 32(3),295-298.doi:10.1177/0265532214565385

Lee, Y-W. (2015) Diagnosing diagnostic language assessment. *Language Testing, 32(3),* 299-316.doi: 10.1177/0265532214565387

Lee, Y-W., & Sawaki, Y. (2009). Cognitive Diagnosis and Q-Matrices in Language Assessment. *Language Assessment Quarterly, 6 (3),* 169-171. doi: 10.1080/154343300903079598

Lee. Y-W., & Sawaki, Y. (2009). Cognitive Diagnosis Approaches to Language Assessment: An Overview. *Language Assessment Quarterly, 6 (3),* 172-189. doi: 10.1080/15434300902985108

Lee, Y-W., & Sawaki, Y. (2009). Application of Three Cognitive Diagnosis Models to ESL Reading and Listening Assessments. *Language Assessment Quarterly, 6 (3),* 239-263. doi: 10.1080/154343300903079562

Lee, Y-W., & Sawaki, Y. (2010). Cognitive Diagnosis and Q-Matrices in Language Assessment:: The Authors Respond. *Language Assessment Quarterly, 7 (1),* 108-112. doi: 10.1080/154343300903559076

Lin, M. (2006). *The effects of note-taking, memory and rate of presentation on EFL learners' listening comprehension.* Unpublished doctoral dissertation, La Sierra University, California.

Lockwood, J. (2013). The Diagnostic English Language Tracking Assessment (DELTA) writing project: a case for post-entry assessment policies and practices in Hong Kong Universities. *Papers in Language Testing and Assessment.* 2 (1). 30-49. Downloaded on 28/10/2016. 10.00 am from http://www.altaanz.org/uploads/5/9/0/8/5908292/3_lockwood.pdf

Luoma, S., & Tarnanen, M.(2003). Creating a self-rating instrument for second language writing: from idea to implementation. *Language Testing*. 20(4), 440-466.

Mansouri, B. (2017). [Review of the book *The Diagnosis of Reading in a Second or Foreign Language,* by J.C. Alderson, E-L. Haapakangas, A. Huhta, L. Nieminem, and R. Ullakonoja]. *Language Assessment Quarterly*, 14(1), 89-92. doi:10.1080./15434303.2016.1261.294

Mason, J. (2007). *Qualitative Researching*. London: Sage Publications.

McKay, P. (2006). *Assessing Young Language Learners*. Cambridge: Cambridge University Press.

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 1-20. doi:10.1177/1362168814567889

McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.

McNamara, D.S., Louwerse, M.M., Cai, Z., & Greaser, A. (2005, January 1). Coh-Metrix version 1.4. Downloaded on 27/09/2016 from http://cohmetrix.memphis.edu.

Moore, R., Adams, E., Dagenais, P., & Caffee, C. (2007). Effects of reverberation and filtering on speech rate judgement. *International Journal of Audiology, 46*, 154-160.

Nation, I.S.P. (2001). *Learning vocabulary in another language.* Cambridge, England: Cambridge University Press.

Oscarrson, M. (1989). Self-assessment of language proficiency: rationale and applications. *Language Testing*. 6 (1), 1-14.

Poehner, M.E., Zhang, J., & Lu, X. (2015).Computerized dynamic assessment (C-DA):Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing, 32(3),* 337-357.doi:10.1177/0265532214560390

Rea-Dickins, P. (2008). Classroom-Based Language Assessment. In *Encyclopedia of language and education* (pp. 2391-2405). Springer US.

Read, J. (2008).Identifying academic language needs through diagnostic assessment. *Journal of English for Academic Purposes, 7,* 180-190

Read, J. (2015). *Assessing English Proficiency for University Study*. Basingstoke: Palgrave McMillan.

Révész, A.,& Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition, 35,* 31-65.

Reiter, R. (1987). A Theory of Diagnosis from first principles. Artificial Intelligence, 32, 57-95.

Richards, J.C. (1983). Listening comprehension: approach, design, procedure. *Tesol Quarterly*, 17 (2), 219-240.

Rosenhouse, J., Haik, L., & Kishon Rabin, L. (2006). Speech perception in adverse listening conditions, in Arabic-Hebrew bilinguals. *International Journal of Bilingualism*, 10(2),119-135.

Rost, M. (2011). *Teaching and Researching Listening* (2nd). Edinburgh Gate: Pearson Education Limited.

Sadighi, F., & Zare, S. (2006). Is listening comprehension influenced by the background knowledge of the learners? A case study of Iranian EFL learners. *The Linguistics Journal*, 1(3), 110-126

Sawaki, Y., Kim, H-J., & Gentile, C.(2009). Q-Matrix Construction: Defining the Link Between Constructs and Test Items in Large-Scale Reading and Listening Comprehension Assessments. *Language Assessment Quarterly,6:3,*190-209.

Sawaki, Y., Quinlan, T., & Lee, Y-W. (2013). Understanding Learner Strengths and Weaknesses: Assessing Performance on an Integrated Writing Task. *Language Assessment Quarterly, 10(1),* 73-95. doi:10.1080/15434303.2011.633305

Sheerin, S. (1987). Listening comprehension: teaching or testing? *ELT Journal*, 41(2), 126-31

Siegel, J. 2015. 'Research Involving L2 listening instruction and instructors'. *ELT Journal* (2015) 69 (3): 323-326.doi: 10.1093/elt/ccv014

Smith, B. (2001). Arabic Speakers. In M. Swan & B. Smith (Eds.) *Learner English, A teacher's guide to interference and other problems*(pp.195-213). Cambridge: Cambridge University Press.

Somekh, B. (1995). The contribution of action research to development in social endeavours: a position paper on action research methodology. *British Education Research Journal*, 21(3), 339-555.

Stine, E.L., Wingfield, A., & Poon, L.W. (1986), How much and how fast: Rapid processing of spoken languages in later adulthood. *Psychology and Aging, 1*(4), 303-311.

Tauroza, S. (2001). Second language lecture comprehension in naturalistic controlled conditions. In J. Flowerdew & M. Peacock (Eds). *Research Perspectives on English for academic purposes* (pp. 360-374). Cambridge: Cambridge University Press.

Turner, C. E., & Purpura, J. E. (2015). Learning-oriented assessment in the classroom. *Handbook of second language assessment. Berlin, Germany/Boston, MA: DeGruyter Mouton*.

Wilson, M. (2003). Discovery listening – improving perceptual processing. *ELT Journal*, 57 (4), 335-43

Unaldi, I. (2014). Self and teacher assessment as predictors of proficiency levels of Turkish EFL learners. *Assessment & Evaluation in Higher Education*, doi: 10.1080/02602938.2014.980223

Urmston, A., Raquel, M., & Tsang, C. (2013). Diagnostic Testing of Hong Kong Tertiary Students' English Language Proficiency: The Development and Validation of DELTA. *Hong Kong Journal of Applied Linguistics,* 14 (2), 60-82.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40, 191-210.

Vandergrift, L., & Goh, C. M. (2012). *Teaching and learning second language listening: Metacognition in action.* New York: Routledge.

Voght, K., & Tsagari, D. (2014). Assessment Literacy of Foreign Language Teachers: Findings of a European Study. *Language Assessment Quarterly,* 11 (4), 374-402.

Vygotsky, L.S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Yin, M. (2008). Diagnosing Foreign Language Proficiency: The interface between learning and assessment. *Language Assessment Quarterly*. 5(1), 77-81

Yin, M., Sims, J., & Cothran, D. (2012). Scratching Where They Itch: Evaluation of Feedback on a Diagnostic English Grammar Test for Taiwanese Students. *Language Assessment Quarterly, 9 (1),* 78-104. doi: 10.1080/15434303.2010.516043

Ying-hui, H. (2006). An investigation into the task features affecting EFL listening comprehension test performance. *The Asian EFL Journal Quarterly*, 8(2), 33-54.

Wingfield, A. (2000). Speech perception and the comprehension of spoken language in adult aging. In D. Park & N.Schwartz (Eds.) *Cognitive Aging: A Primer.* (pp. 175-195). Philadelphia, PA: Psychology Press.

Zhao, Y. (1997). The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics*, 18(1), 49-68.

Zielinski, B.W.(2008). *The Listener: No longer the silent partner in reduced intelligibility.* Science Direct, System 36, 69-84.

Appendix A - GSoE Research Ethics Form

It is important for members of the Graduate School of Education, as a community of researchers, to consider the ethical issues that arise, or may arise, in any research they propose to conduct. Increasingly, we are also accountable to external bodies to demonstrate that research proposals have had a degree of scrutiny. *This form must therefore be completed for each piece of research carried out by members of the School, both staff and students.*

The GSoE's process is designed to be supportive and educative. If you are preparing to submit a research proposal, you need to do the following:

1. **Arrange a meeting with a fellow researcher**
   The purpose of the meeting is to discuss ethical aspects of your proposed research, so you need to meet with someone with relevant research experience. A list of prompts for your discussion is given below. Not all these headings will be relevant for any particular proposal.
2. **Complete the form on the back of this sheet**
   The form is designed to act as a record of your discussion and any decisions you make.
3. **Upload a copy of this form and any other documents (e.g. information sheets, consent forms) to the online ethics tool at :** https://dbms.ilrt.bris.ac.uk/red/ethics-online-tool/applications**.**

   Please note: Following the upload you will need to answer ALL the questions on the ethics online survey and submit for approval by your supervisor (see the flowchart and user guides on the GSoE Ethics Homepage).

If you have any questions or queries, please contact the ethics co-ordinators at: gsoe-ethics@bristol.ac.uk

Please ensure that you allow time before any submission deadlines to complete this process.

Prompts for discussion

You are invited to consider the issues highlighted below and note any decisions made. You may wish to refer to relevant published ethical guidelines to prepare for your meeting. See http://www.bristol.ac.uk/education/research/networks/ethicnet for links to several such sets of guidelines.

1. Researcher access/ exit
2. Information given to participants
3. Participants right of withdrawal
4. Informed consent
5. Complaints procedure
6. Safety and well-being of participants/ researchers

7. Anonymity/ confidentiality
8. Data collection
9. Data analysis
10. Data storage
11. Data Protection Act
12. Feedback
13. Responsibilities to colleagues/ academic community
14. Reporting of research

Be aware that ethical responsibility continues throughout the research process. If further issues arise as your research progresses, it may be appropriate to cycle again through the above process.

Name(s): *Sohaib Meeran Sandhu*

Proposed research project:

Diagnostic listening Assessment- A Saudi academic listening context

Proposed funder(s): *Not applicable*

Discussant for the ethics meeting: *Muhibulah-Abdur-Rahman, EdD (Bristol)*

Name of supervisor: *Talia Isaacs*

Has your supervisor seen this submitted draft of your ethics application? Y/~~N~~

Please include an outline of the project or append a short (1 page) summary:

The main objective of this research is to add knowledge to the area of diagnostic assessment, especially diagnostic listening assessment. It seeks to explore and examine how practical, feasible and implementable diagnostic listening assessment is in resolving learner listening problems and thus be able to help them progress and learn how to listen, something that Field (2008) believes there is a lack of, where the listening teaching process largely follows a testing approach. It is also hoped that the findings will help to create a bridge between what is largely a theoretical idea, to what is possible in practice. The research study will be conducted over 1 academic year (2 semesters) and will involve diagnosing learner problems relating to listening, with the aid of discussions, self-assessment forms, and diagnostic tests. They will be provided with tasks to do to help alleviate perceived problems, and then be involved in a review of their progress. Teachers will be involved, and their views on the diagnostic process will also be drawn upon.. The research study will also be open to any learner who feels that he needs help to improve listening ability.

Ethical issues discussed and decisions taken (see list of prompts overleaf):

1. **Researcher access/ exit**

I have been informally told (via email) by the director that I should put in writing what I need but there should be no problem with access I believe to lab facilities and students.

The target participants will be teachers and students, with varying levels of involvement.

Teacher Recruitment:

- Teachers will be invited by email to a workshop presentation of the research project. They will then be asked if they can volunteer.

Possible sample script for an email to teachers:

"Dear colleague, you are invited to a presentation which will outline a research project that I am proposing to conduct this academic year as part of my doctoral studies. The research project aims and objectives will be outlined, the possible benefits of the research for learners and the programme generally, and how teachers could also possibly benefit by being involved in the research process. Attendance to the workshop will not imply any kind of commitment to the research project in any way; however, I will use this workshop as an opportunity to request participation in the research project".

Student recruitment will be approached in four ways:

- Teacher recommendation: Where the teacher brings the learner to the Diagnostic language clinic and will have a case file opened for him (There are only male students). However, student participation will be voluntary.
- There will be general advertising via noticeboards, inviting those who wish to improve their English to come along to the clinic. The notice will be written in Arabic too.
- An email may be sent out to all students informing them of the project and the possibility of their voluntary involvement.
- A circular in Arabic will be distributed to learners in Arabic, possibly going to each classroom, outlining briefly the benefits of receiving a free consultation on their listening abilities, and possibility of being able to improve their language.

2. Information given to participants

- Information sheet in English and with an Arabic translation
- Consent form in English and with an Arabic translation

Both teachers and students will be provided with an information sheet detailing what the research project is about. For students, this will be provided in Arabic. In addition to this, a consent form will be provided which will be signed by both the research participant as well as the researcher. A copy of the information sheet and consent form will be given to the research participants.

3. Participants right of withdrawal

Participant withdrawal without any consequence will be detailed on the information sheet and made clear at the very start. Withdrawal will have no impact on grades or have any adverse effect on the research participant in any way.

4. **Informed consent**

Participants will be asked to read an information sheet and sign a letter of consent.

5. **Complaints procedure**

If research participants wish to complain about any aspects of the research process, they will be able to speak to an appointed independent person. The researchers' supervisor will be named as the person to complain to in the event of any complaint.

6. **Safety and well-being of participants/ researchers**

As the research will be conducted in the computer labs all appropriate precautions will be taken, and existing safety regulations implemented.

7. **Anonymity/ confidentiality**

The letter of consent will be written in English and will be translated into Arabic. Anonymity and confidentiality of data will be made explicit and learners will have the right to seek re-assurance from the researcher. When writing up the research report, pseudonyms will be used instead of actual research participant names.

8. **Data collection**

Basic student and teacher data will be used to help cross reference or link any information collected. Data is proposed to be collected via the following methods:

- An initial interview which may be recorded
- An initial self-assessment form
- Diagnostic test(s) and tasks
- Teachers may be interviewed for their views on the learner
- Students will be interviewed later to get their views on the process of learning
- A personal log or journal on the experience of diagnosing learners by the researcher

9. **Data analysis**

Data is expected to be analysed using a variety of qualitative and quantitative methods. Data will not be analysed in a way that somehow negatively impacts on the learners who participate. Data will be reported so as not to identify participants personally.

10. Data storage

Data will be stored in a secure password protected drive which will be in possession of the researcher. Any paper-based data will be kept in secure premises out of the reach of anyone. This is likely to be the residence of the researcher.

11. Data Protection Act

Data protection Act regulations that exist in both in the United Kingdom and Saudi Arabia will be enacted. In addition, Taibah University's (the research location) own regulations on these matters will also be taken into account with regards to data collection. Any response data that is viewed or viewable by others, will be collated in a way that does not link the research subject directly to the response data.

12. Feedback

Feedback will be dealt with sensitively and in confidence, ensuring that there are no negative effects on research subject. Feedback will be based on the self-assessment forms provided by the learners, their responses to listening tasks, and the tasks they need to do to help them improve their perceived weaknesses where relevant.

13. Responsibilities to colleagues/ academic community

No one will be coerced or pressurised into referring learners to the listening diagnostician. Learners will be asked to volunteer for the research.

One of the problems where learners' own teachers are involved in the process, is the issue of power relations. Where teachers may wish to refer a learner to the diagnostic language centre, learners may comply in order to make their teacher happy. Procedures will be put into place that do not allow for the coercion of learners into the research process.

14. **Reporting of research**

All participants will be anonymised through the use of pseudonyms, unless they request otherwise.

If you feel you need to discuss any issue further, or to highlight difficulties, please contact the GSoE's ethics co-ordinators who will suggest possible ways forward.

Signed: *Sohaib Meeran Sandhu* (Researcher) Signed: *Dr. Talia Isaacs* (Discussant)

Date: 14th July 2015

Appendix B - Listening (diagnostic) review sheet

الاسم: Name:	الرقم الجامعي: University ID Number:
رقم وعنوان الوحدة: Chapter/Unit number and name:	اسم مقطع الاستماع: Listening Text name:
الرجاء الاجابة على السؤال التالي باللغة العربية. *هل تعلم أي شيء عن موضوع مقطع الاستماع؟ اذا كانت الاجابة بنعم فالرجاء كتابة كل ما تستطيع كتابته. بعدها حاول كتابة ما تعتقد بأنك ستستمع له في المقطع.* Please answer this question in Arabic. Do you know anything about this topic? If yes, could you write as much as you can? Then, try to write down what you think you'll hear in the passage.	
الرجاء الاجابة على السؤال التالي باللغة العربية. *ماهي الكلمات الانجليزية التي تتوقع سماعها في المقطع؟ هل تستطيع التخمين؟ ان لم تكن تعرف الكلمات المتوقع سماعها باللغة الانجليزية بإمكانك كتابتها باللغة العربية.* Please answer this question in Arabic What kinds of English words do you think they might hear? Can you guess them? If you don't know the words in English, write them in Arabic.	
الآن أستمع للمقطع الصوتي Now listen to the audio	
هل بإمكانك كتابة النقاط الرئيسية للمقطع الصوتي الذي استمعت له للتو؟ بإمكانك طلب اعادة تشغيل المقطع من المعلم. بإمكانك الكتابة في الجهة	

الأخرى من الورقة في حالة الحاجة. Can you summarise in Arabic the main points of what you heard? Ask the teacher to play the audio again if required. Write on the other side if required.	
تي؟ ضع دائرة حول الاجابة المناسبة. Did you understand the general idea?	نعم – لا YES / NO
هل وجدت موضوع مقطع الاستماع مثير وممتع؟ ضع دائرة حول الاجابة المناسبة. Did you find the topic interesting? Circle one.	نعم – لا – جزء منه – لست متأكدا Yes / No / Some of it / I'm not sure
ما مقدار استيعابك لمقطع الاستماع؟ ضع علامة اكس ( ) على الخط لتبين مدى استيعابك. How much of the audio did you understand? Put a CROSS (X) on the line that represents how much you felt you understood.	☹----------|----------|----------|----------☺ 25% 50% 75%
سرعة المحادثة في المقطع الصوتي (ضع دائرة حولة أحد الخيارات التالية) Pace or speed of recording (Circle one of the choices on the right)	1(أعتقد بأنها كانت سريعة جدا) 2(أعتقد بأنها كانت سريعة نوعا ما) 3(أعتقد بأن السرعة كانت مناسبة) 4(أعتقد بأنها كانت بطيئة نوعا ما) 5(أعتقد بأنها كانت بطيئة جدا) 1.(I thought it was very fast) 2.(I thought it was slightly fast) 3.(It was at just the right speed) 4.(It was slightly slow) 5.(It was too slow)
هل كان في المقطع الصوتي كلمات جديدة عليك تسمعها للمرة الأولى؟ Did you hear any words that you have	نعم - لا YES / NO

not heard before?	
هل كان في المقطع الصوتي كلمات سبق أن سمعتها لكنك لم تتعرف على معناها؟ Were there any words that you recognized but could not remember what they meant?	نعم – لا YES / NO
هل واجهت أي مشاكل أخرى في مقطع الاستماع؟ الرجاء كتابتها. Did you have any other problems? Please tell us here.	
استمع إلى الأفكارا الرئيسيةListen for Main Ideas ضع دائره حول أفضل إجابه. الأسئله كانت سهله متوسطه صعبه Circle the best answer. The questions were Easy, Okay, Difficult	**Listen for Details استمع إلى التفاصيل** ضع دائره حول أفضل إجابه. الأسئله كانت سهله متوسطه صعبه Circle the best answer. The questions were Easy, Okay, Difficult
أنا أفهم المفترض بي أن أفعله؟ نعم /لا/ أحيانا؟ I understand what I was supposed to do? YES / NO / Sometimes	I undertand what I was supposed to do? YES / NO / Sometimes
فهمت كل الكلمات في الاسئلة والإجابات؟ نعم/لا/ أحيانا؟ I understood all of the words in the questions and answers? YES / NO / Sometimes	فهمت كل الكلمات في الاسئلة والإجابات؟ نعم/لا/ أحيانا؟ I understood all of the words in the questions and answers? YES / NO / Sometimes
هل كانت كل إجاباتك صحيحه؟ نعم \ لا Did you get all of the answers right? YES / NO	هل كانت كل إجاباتك صحيحه؟ نعم \ لا Did you get all of the answers right? YES / NO
هل واجهت أي مشاكل في الحلول؟ الرجاء الشرح Were there any other problems that you had? Please explain For example Sound/ Accent/ Anything else أي شيء آخر/ لهجة/الصوت :فمثلا	هل واجهت أي مشاكل في الحلول؟ الرجاء الشرح Were there any other problems that you had? Please explain For example Sound/ Accent/ Anything else أي شيء آخر/ لهجة/الصوت :فمثلا

Appendix C – Taxonomy of Listening Skills, Richards (1983, p.228-229)

Micro-skills: Conversational Listening

1. Ability to retain chunks of language of different lengths for short periods
2. Ability to discriminate among the distinct sounds of the target language
3. Ability to recognise the stress patterns of words
4. Ability to recognise the rhythmic structure of English
5. Ability to recognise the functions of stress and intonation to signal the information structure of utterances
6. Ability to identify words in stressed and unstressed positions
7. Ability to recognise reduced forms of words
8. Ability to distinguish word boundaries
9. Ability to recognise typical word order patterns in the target language
10. Ability to recognise vocabulary used in core conversational topics
11. Ability to detect key words (i.e. those which identify topics and propositions)
12. Ability to guess the meanings of words from the contexts in which they occur
13. Ability to recognise grammatical word classes (parts of speech)
14. Ability to recognise major syntactic patterns and devices
15. Ability to recognise cohesive devices in spoken discourse
16. Ability to recognise elliptical forms of grammatical units and sentences
17. Ability to detect sentence constituents
18. Ability to distinguish between major and minor constituents
19. Ability to detect meanings expressed in differing grammatical forms/sentence types (i.e., that a particular meaning may be expressed in different ways)
20. Ability to recognise the communicative functions of utterances, according to situations, participants goals
21. Ability to reconstruct or infer situations, goals, participants, procedures
22. Ability to use real world knowledge and experience to work out purposes, goals, settings, procedures
23. Ability to predict outcomes from events described
24. Ability to infer links and connections between events
25. Ability to deduce causes and effects from events
26. Ability to distinguish between literal and implied meanings
27. Ability to identify and reconstruct topics and coherent structure from ongoing discourse involving two or more speakers
28. Ability to recognise markers of coherence in discourse, and to detect such relations as main idea, supporting idea, given information, new information, generalisation, exemplification
29. Ability to process speech at different rates
30. Ability to process speech containing pauses, errors, corrections
31. Ability to make use of facial, paralinguistic, and other clues to work out meanings
32. Ability to adjust listening strategies to different kinds of listener purposes or goals
33. Ability to signal comprehension or lack of comprehension, verbally and non-verbally

Appendix D - Taxonomy of Listening Skills, Richards (1983, p.229-230)

Micro-skills: Academic Listening (Listening to Lectures)

1. Ability to identify purpose and scope of lecture
2. Ability to identify topic of lecture and follow topic development
3. Ability to identify relationships among units within discourse (e.g., major ideas, generalisations, hypothesis, supporting ideas, examples)
4. Ability to identify role of discourse markers in signalling structure of a lecture (e.g., conjunctions, adverbs, gambits, routines)
5. Ability to infer relationships (e.g., cause, effect, conclusion)
6. Ability to recognise key lexical items related to subject/topic
7. Ability to deduce meanings of words from context
8. Ability to recognise markers of cohesion
9. Ability to recognise function of intonation to signal information structure (e.g., pitch, volume, pace, key)
10. Ability to detect attitude of speaker toward subject matter
11. Ability to follow different modes of lecturing: spoken, audio, audio-visual
12. Ability to follow lecture despite differences in accent and speed
13. Familiarity with different styles of lecturing: formal, conversational, read, unplanned
14. Familiarity with different registers: written vs colloquial
15. Ability to recognise irrelevant matter: jokes, digressions, meanderings
16. Ability to recognise function of non-verbal cues as markers of emphasis and attitude
17. Knowledge of classroom conventions (e.g., turn taking, clarification requests)
18. Ability to recognise instructional/learner tasks (e.g., warnings, suggestions, recommendations, advice, instructions)

Appendix E – Fields Decoding processes

(Field, 2008, Appendix 1, p.336-337)

Decoding processes

1. Phoneme Level
1.1 Phoneme recognition in a range of contexts
1.2 Discriminating consonants
1.3 Discriminating vowels
1.4 Recognising consonant clusters
1.5 Extrapolating spellings from sounds
2. Syllable Level
2.1 Recognising syllable structure
2.2 Recognising syllable stress
2.3 Treating stressed syllables as more reliable
2.4 Using stressed syllables as access codes
2.5 Using weak syllables to locate function words
3. Word level
3.1 Lexical segmentation
3.1.1 Rhythm-based strategies
3.1.2 Using prefixes and suffixes as boundary markers
3.1.3 Using fixed stress (where appropriate)
3.2 Recognising variant forms of words
3.2.1 Allowing for cliticisation
3.2.2 Allowing for resyllabification
3.2.3 Recognising weak forms of function words
3.2.4 Recognising assimilated words
3.2.5 Allowing for elision
3.2.6 Recognising reduced words within intonation groups
3.3 Recognising complete formulaic chunks
3.4 Using awareness of word frequency
3.5 Current activation
3.6 Spreading activation (word networks in the mind)
3.7 Distinguishing known and unknown words
3.8 Dealing with unknown words: infer – generalise – ignore
3.9 Automatic lexical access
4. Syntactic parsing
4.1 Building syntactic structures during pauses and fillers
4.2 Using planning pauses to demarcate syntactic structures
4.3 Distinguishing planning and hesitation pauses
4.4 Using intonation groups to demarcate syntactic structures
4.5 Building a syntactic structure online
4.5.1 Testing hypotheses

4.5.2 Using probability
4.5.3 Recognising syntactic chunks
4.5.4 Recognising the sentence pattern associated with the verb
4.5.5 Recognising primary L2 cues to syntactic organisation
4.6 Understanding functional language
4.7 Drawing inferences based on syntax
5. Intonation level
5.1 Relating intonation groups to syntactic structure
5.2 Forming and testing decoding hypotheses as an intonation group proceeds
5.3 Identifying focally stressed syllables
5.4 Treating focally stressed syllables as central to the message
5.5 Recognising recurrent intonation-group chunks
5.6 Guessing words of low prominence in the intonation group
6. Normalisation to speaker voices
6.1 Allowing for voice variation
6.2 Setting baseline for loudness, pitch level, speech rate
6.3 Drawing on an accent repertoire

Appendix F – Fields Meaning-building processes

(Field, J., 2008, Appendix B, p.338-339)

Meaning-building processes

1. Word meaning
1.1 Narrowing word sense to fit context
1.2 Dealing with word ambiguity
1.3 Inferring meaning of unknown words
2. Syntactic meaning
2.1 Relating syntax to context
2.2 Interpreting speaker's functional intentions
2.3 Forming inferences from syntactic information
3. Intonation meaning
3.1 Recognising given/new relationships
3.2 Distinguishing given/new and contrastive and emphatic stress
3.3 Relating contrastive and emphatic focal stress to context
3.4 Recognising finality
3.5 Recognising the end of a speaker turn
3.6 Using intonation to identify questions in statement form
3.7 Distinguishing a confirmation request from a more open question
3.8 Distinguishing echoes and challenges
3.9 Distinguishing neural – emotive – withdrawn intonations
4. Using contextual knowledge
4.1 World knowledge
4.2 Topic knowledge
4.3 Speaker knowledge
4.4 Knowledge of situation
4.5 Knowledge of setting
5. Using schematic knowledge (including scripts)
5.1 Predicting what will be said
5.2 Triggering spreading activation
5.3 Inferring what the speaker has not expressed
5.4 Allowing for culturally determined schemas
6. Context/co-text and meaning
6.1 Using context and co-text to narrow down word meaning
6.2 Using context and co-text to infer pragmatic meaning
6.3 Using context and co-text to infer word meaning
7. Using inference
7.1 Inferring information the speaker has left unsaid
7.2 Inferring connections between pieces of information that were not made explicitly
8. Making reference connections
8.1 Carrying forward current topics

8.2 Dealing with imprecise reference
9. Interpreting the utterance
9.1 Interpreting speaker language
9.2 Deep processing
10. Selecting information
10.1 Considering relevance
10.2 Considering redundancy: addition versus repetition
10.3 Dealing with incoherence
11. Integrating information
11.1 Connecting new information to previous
11.1.1 Recognising locally connecting linkers
11.1.2 Recognising ‘signpost’ linkers
11.1.3 Recognising links not marked by linkers
11.2 Monitoring for consistency
11.3 Structuring for discourse
11.3.1 Recognising topics and sub-topics
11.3.2 Using formal schemas
12. Forming and checking provisional discourse representations
12.1 Forming the basis for a discourse representation
12.2 Accepting an indeterminate representation
12.3 Checking, revising and upgrading a discourse representation

Appendix G - Research Information Sheet and Consent form (English version)

Research Information Sheet (Students)

Researcher name: Sohaib Meeran Sandhu

Supervisor name: Dr.Talia Isaacs

Aims of the Project:

To assess problems in English language listening, and to help learners improve their listening ability.

You are invited to take part in a research project. This project will help us to understand what listening problems learners have, and what we can do to help them.

Process

- You will be asked to listen to some English language passages and do some listening tasks.
- You will be asked about the difficulties you have in listening and this will be recorded.
- You will be asked to evaluate your own listening ability which will mean filling out forms
- We will try to analyse the problems and suggest ways in which you can improve your listening.
- We will then ask you to do more tests to see if there has been any improvement and get your opinion on this process.

How often:

- We require your participation for about 30-60 minutes a week for about 4 weeks.

Potential benefits

- It is hoped that whatever advice and tasks we give you should benefit your studies directly.

Your information and data

- All information that you provide will be strictly confidential. At no time will we give out information to anyone other than to research staff or the participating teacher.
- The results of this study may be presented at professional meetings or published in a professional journal, but your name and identity will not be revealed.
- Any data collected will be stored in a secure area.

Your rights:

You have the right to withdraw at any time you wish. Your withdrawal will have no effect on your grades and you have nothing to worry about. You will not be penalised in any way by withdrawing from the research project

Complaints procedure

If you are unhappy in any way about any aspects of the research, you can contact Dr. Talia Isaacs on 00441173314312 or email her at talia.isaacs@bristol.ac.uk

Student Consent Form

Researcher name: Mr. Sohaib Meeran Sandhu

Supervisor name: Dr.Talia Isaacs

Project Title:

Exploring diagnostic listening assessment in an academic environment: A Saudi context

Please Initial Box

1. I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving reason. ☐

3. I agree to take part in the above study. ☐

Student ID of Participant	Date	Signature
Name of Researcher	Date	Signature

Appendix H - (Student Information Sheet and consent form – Arabic version)

(Student Information Sheet and consent form – Arabic version)

نموذج بيانات بحث (طلاب)

اسم الباحث : صهيب ميران صاندو

اسم المشرف : د. تاليا أيزاكس

أهداف البحث :

تقييم مشاكل الاستماع في اللغة الانجليزية ومساعدة المتعلمين على تحسين مهارة الاستماع لديهم.

انت مدعو للمشاركة في مشروع بحث. هذا المشروع سوف يساعدنا على فهم مشاكل الاستماع لدى المتعلمين وما يمكننا القيام به لمساعدتهم.

سيرالعملية:

- سوف يطلب منك الاستماع إلى بعض المقاطع باللغة الإنجليزية والقيام ببعض تدريبات الاستماع.
- سوف تسأل عن صعوبات الاستماع لديك وسوف يتم تسجيل ذلك.
- سوف يطلب منك تقييم مهارة الاستماع لديك وهذا يتطلب ملئ استمارات.
- سوف نحاول تحليل لمشاكل واقتراح الطرق التي تمكنك من تحسين الاستماع لديك.
- سوف نطلب منك بعد ذلك أن تقوم بالمزيد من الاختبارات لمعرفة ما إذا كان هناك أي تحسن والحصول على رأيك في هذا البحث.

مدة المشاركة:

- نحن بحاجة إلى مشاركتك لمدة تتراوح ما بين 30 إلى 60 دقيقة في الأسبوع ولمدة 4 أسابيع.

الفوائدالمحتملة:

- من المؤمل أن تستفيد في دراستك استفادة مباشرة من النصائح والمهام التي تقدم لك.

المعلومات والبيانات الخاصة بك:

- ستكون جميع المعلومات التي تقدمها سرية للغاية. لن نسلم هذه المعلومات إلى أي أشخاص سوى القائمين على البحث أو المدرس المشارك.
- من الممكن أن يتم تقديم نتائج هذه الدراسة في اللقاءات المهنية أو نشرها في مجلة علمية، و لكن لن يتم الكشف عن اسمك أو هويتك.
- جميع البيانات التي سيتم جمعها ستحفظ في مكان آمن.

حقوقك :

لديك الحق في الانسحاب في أي وقت تشاء. لن يكون لانسحابك أي تأثيرسلبي.

إجراءات الشكاوى :

إذا كنت غير راض بأي شكل من الأشكال عن أي جانب من جوانب البحث، يمكن الاتصال المشرفة المذكورة سابقاً، الدكتورة تاليا أيزاكس، وذلك على الرقم: 00441173314312 أوعلى بريدها الالكتروني:

talia.isaacs@bristol.ac.uk

نموذج موافقة طالب

اسم الباحث :صهيب ميران صاندو

اسم المشرف : د. تاليا أيزاكس

عنوان البحث :

Exploring diagnostic listening assessment in an academic environment: A Saudi context

	ضع علامة في المربع
1. أؤكد أنني قرأت وفهمت نموذج بيانات الدراسة المذكورة أعلاه، وأتيحت لي الفرصة للإستفسار.	☐
2. أفهم أن مشاركتي طوعية ولي الحرية في الانسحاب في أي وقت، دون ذكر السبب.	☐
3. أوافق على المشاركة في الدراسة المذكورة أعلاه.	☐

_____ _____ _____

الرقم الجامعي للطالب المشارك التاريخ التوقيع

_____ _____ _____

اسم الباحث التاريخ التوقيع

Appendix I - Student Background Questionnaire (English version)

Student Background Questionnaire (English version)

1. Student ID number _____
2. Birthplace (City, Country): _____
3. Age: _____
4. Is your hearing normal as far as you know? YES / NO
5. First languages (s) from birth:_____
6. Mothers first language: _____
7. Father first language: _____
8. How old were you when you first started to learn English? _____
9. Did you study English at school? YES / NO
    a. If yes, at which stage of school? (please tick √)
        i. Primary _____
        ii. Secondary _____
        iii. High School _____
10. Have you studied English at any language centre or have you had English tuition?

    YES /NO

11. Do you speak English at home? **Circle** one answer:

    - Yes
    - No
    - Sometimes

12. Do you speak English with anyone else for example friends or others? If YES, please write down who you speak to below:

_____

13. Approximately what percentage of the time do you speak in the English Language (rather than in any other language) in your daily life?

    0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

14. Approximately what percentage of the time do you listen to the English Language in the media (rather than in any other language) in your daily life?

    0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

15. Have you ever lived in or visited a country where you had to speak English? YES /NO. If you answered YES, please answer the questions below. If you answered NO, go to the next question.

    a. Name the country or countries in which you either lived or visited where you had to speak English and how long you were there for:

    Name of place How long were you there?

16. Have you ever listened to English in different accents? YES / NO

17. If yes, which accents have you heard? Please write them below, and say in what situation (for example when watching a movie, meeting someone, my teacher etc.)

    Accent Situation

18. When listening, which accent or accents do you prefer? Please **Circle** any of the answers below. You can choose as many answers as you wish.

    - American

- British
- Saudi
- Egyptian
- Jordanian
- Syrian
- Lebanese
- Indian
- Pakistani
- Other (please write these down): _____
- I like and prefer all accents

19. When listening to a recording or when speaking to someone, which of the following makes listening or understanding difficult? Please **Circle** all of the relevant answers:

    - Accent
    - Speed of speaking
    - Vocabulary I've never heard before
    - Vocabulary I think I've heard before, but which I can't remember the meaning of.
    - Grammar: In other words, I get confused when the word might be used like a verb, a noun or an adjective.
    - Sometimes, I'm not sure when words are combined with other words.
    - To be honest, I'm not really sure why I don't understand the listening

20. How good do you think your English is? **Circle** one answer:

    - Excellent
    - Very Good
    - Good
    - Okay
    - Bad
    - Very bad
    - I'm not sure

21. How do you rate your English language skills? Please put the numbers 1, 2, 3 or 4 against each skill, with 1 being the best skill and 4 being the worst.
    - Reading ___
    - Writing ___
    - listening ___
    - speaking ___

22. Have you ever taken an English proficiency or any other exam like TOEFL, IELTS, PET, KET etc)? YES /NO

23. If YES, what was your score? _____

Appendix J - Student Background Questionnaire (Arabic version)

Student Background Questionnaire (Arabic version)

استبيان عن خلفية الطالب

1. الرقم الجامعي: _____
2. مكان الولادة (المدينة والبلد) : _____
3. العمر : _____
4. هل حاسة السمع لديك سليمة – على حد علمك ؟ نعم / لا
5. اللغة (اللغات) الأم – منذ الولادة : _____
6. اللغة الأم للأم : _____
7. اللغة الأم للأب : _____
8. كم كان عمرك عندما بدأت تعلم اللغة الانجليزية لأول مرة؟ _____
9. هل درست اللغة الانجليزية بالمدرسة ؟ نعم / لا

أ. إذا كانت الإجابة بنعم ضع علامة (√) أمام الاختيار المناسب

(i) المرحلة الإبتدائية _____

(ii) المرحلة المتوسطة _____

(iii) المرحلة الثانوية _____

10. هل درست اللغة الانجليزية بمعهد لغة أولدى مدرس خصوصي ؟ نعم / لا
11. هل تتحدث اللغة الانجليزية في المنزل ؟ ضع دائرة حول إجابة واحدة :

- نعم

- لا
- أحيانا

12. هل تتحدث اللغة الانجليزية مع أشخاص آخرين مثل الأصدقاء, المعلم وغيرهم؟ إذا كانت الإجابة بنعم فأذكر علاقتك بالأشخاص الذين تتحدث معهم في الفراغ أدناه:

_____

13. ما هي النسبة التقريبية لوقت تحدثك باللغة الانجليزية (دون غيرها) في حياتك اليومية؟

%10-0	%20-11	%30-21	%40-31	%50-41	%60-51	%70-61	%80-71	%90-81	%100-91

14. ما هي النسبة التقريبية لوقت استماعك للغة الانجليزية (دون غيرها) في وسائل الاعلام -على سبيل المثال يوتيوب و التلفاز- في حياتك اليومية؟

%10-0	%20-11	%30-21	%40-31	%50-41	%60-51	%70-61	%80-71	%90-81	%100-91

15. هل سبق وأن أقمت في بلد أو زرت بلدا توجب عليك فيه التحدث باللغة الانجليزية ؟ نعم / لا
إذا كانت الإجابة **بنعم**، فأجب فضلا عن السؤال أدناه وإذا كانت الإجابة **بلا**، فاذهب الى السؤال التالي.

أ. اذكر اسم البلد او البلدان التي توجب عليك فيها التحدث باللغة الانجليزية والفترة التي قضيتها بها :

اسم المكان	**مدة الإقامة**

16. هل سبق لك الاستماع للغة الانجليزية بلهجات مختلفة ؟ نعم / لا

17. إذا كانت الإجابة **بنعم**، فماهي اللهجات التي استمعت إليها؟ الرجاء كتابة اللهجات بالأسفل وذكر الحالة التي كان فيها ذلك (مثال: مشاهدة فلم أو لقاء شخص أو المدرس إلخ...)

اللهجة	الحالة
.....	.....
.....	.....
.....	.....

(* أنظر إلى قائمة اللهجات في أسفل السؤال 18)

18. أي لهجة أو لهجات إنجليزية تفضل عند الإستماع ؟ ضع دائرة حول الإجابات المناسبة أدناه. يمكنك اختيار أي عدد من اللهجات.

- الأمريكية
- البريطانية
- السعودية
- المصرية
- الشامية (الأردنية / السورية / اللبنانية)
- الهندية / الباكستانية
- أخرى (الرجاء كتابتها) : _____

- أحب وأفضل كل اللهجات

أي مما يلي يجعل الاستماع والفهم صعبا عند الاستماع إلى تسجيل صوتي أو التحدث إلى شخص ما؟ يرجى وضع دائرة حول الإجابات المناسبة:

- اللهجة
- سرعة التحدث
- المفردات التي لم أسمعها من قبل
- المفردات التي أعتقد أنني سمعتها من قبل ولكن لا أذكر معانيها
- القواعد النحوية – أي الشعور بالحيرة عند احتمال استعمال الكلمة كفعل أو اسم أو صفة
- احيانا لا أكون متأكدا عند جمع كلمات مع كلمات أخرى
- لست متأكدا حقا لماذا لا أفهم الإستماع

19.كيف تقيم مستواك في اللغة الانجليزية ؟ اختر إجابة واحدة :

- ممتاز
- جيد جدا
- جيد
- مقبول
- سيء
- سيء جدا
- لست متأكدا

تقييم مهارات اللغة الإنجليزية لديك

20.كيف تقيم مهارات اللغة الانجليزية لديك؟ رتب المهارات التالية حسب مقدرتك اللغوية. ضع الأرقام 1، 2، 3 أو 4 أمام كل مهارة، علما بأن 1 يعني المهارة الأفضل و4 المهارة الأسوأ.

- القراءة _____
- الكتابة _____
- الإستماع _____
- المحادثة _____

21.هل سبق وأن خضعت لاختبار كفاءة لغوية أو أي اختبار آخر مثل توفل (TOEFL)، أيلتس (IELTS)، PET، KET، إلخ؟ **نعم / لا**

22.إذا كانت الإجابة **بنعم**، فأذكر نوع الإختبار والدرجة التي حصلت عليها.

الاختبار	**الدرجة**

Appendix K - Expert Bio Questionnaire: Piloting of research instruments by 'experts'

Expert Bio Questionnaire: Piloting of research instruments by 'experts'

Instruments to be piloted:

- Student information sheet and consent form in **Arabic**
- Student background questionnaire in **Arabic**

Expert code: _____ (to be provided by researcher)

What qualifications do you have?

_____

How many years EFL / ESL teaching experience do you have?

_____

How would you describe your proficiency in the following languages? Please circle one for each language.

English

- Excellent
- Very good
- Good
- Less than good

Arabic

- Excellent
- Very good
- Good
- Less than good

Signature: _____

Date: _____

You can add any notes if you wish to the back of this sheet

Appendix L - Individual profile 2 - Main Idea and details – MOKH01

Individual Profile 2 – Main Idea Tasks MOKH01 – CEFR A2

Lesson	MAIN IDEA **TASK** Difficulty	MAIN IDEA **TASK** Understood what to do?	MAIN IDEA **TASK** Understood Vocabulary	MAIN IDEA **TASK** All Questions Right?	MAIN IDEA **TASK** Any Other Problems
BK2U6L1	EASY	N/A	N/A	-	N/A
BK2U6L2	HARD	-	-	-	LOTS OF INFO ANSWERS SIMILAR
BK2U7L1	EASY	SOMETIMES	SOMETIMES	-	-
BK2U7L2	EASY	Y	Y	Y	ACCENT
BK2U8L1	OKAY	SOMETIMES	SOMETIMES	N 4/6	ACCENT
BK2U8L2	EASY	Y	Y	N 3/4	-
BK2U9L1	OKAY	SOMETIMES	SOMETIMES	N	-
BK2U9L2	OKAY	Y	Y	Y	-
BK3U1L1	-	-	-	-	-
BK3U1L2	EASY	Y	Y	N	-

Individual Profile 2 – Detail Tasks - MOKH01 – CEFR A2

Lesson	LISTEN FOR DETAILS **TASK** Difficulty	LISTEN FOR DETAILS **TASK** Understood what to do?	LISTEN FOR DETAILS **TASK** Understood Vocabulary	LISTEN FOR DETAILS **TASK** All Questions Right?	LISTEN FOR DETAILS **TASK** Any Other Problems
BK2U6L1	EASY	N/A	N/A	N/A	N/A
BK2U6L2	EASY	`-	-	-	-
BK2U7L1	EASY	SOMETIMES	SOMETIMES	-	-
BK2U7L2	EASY	Y	Y	N 7/8	ACCENT
BK2U8L1	OKAY	Y	Y	Y 3/3	ACCENT
BK2U8L2	EASY	Y	N	N 3/4	-
BK2U9L1	EASY	Y	Y	Y	-
BK2U9L2	EASY	Y	Y	Y	-
BK3U1L1	-	-	-	-	-
BK3U1L2	OKAY	Y	N	N	-

Appendix M - Individual profile 3 - Main Idea and details – ABAL02

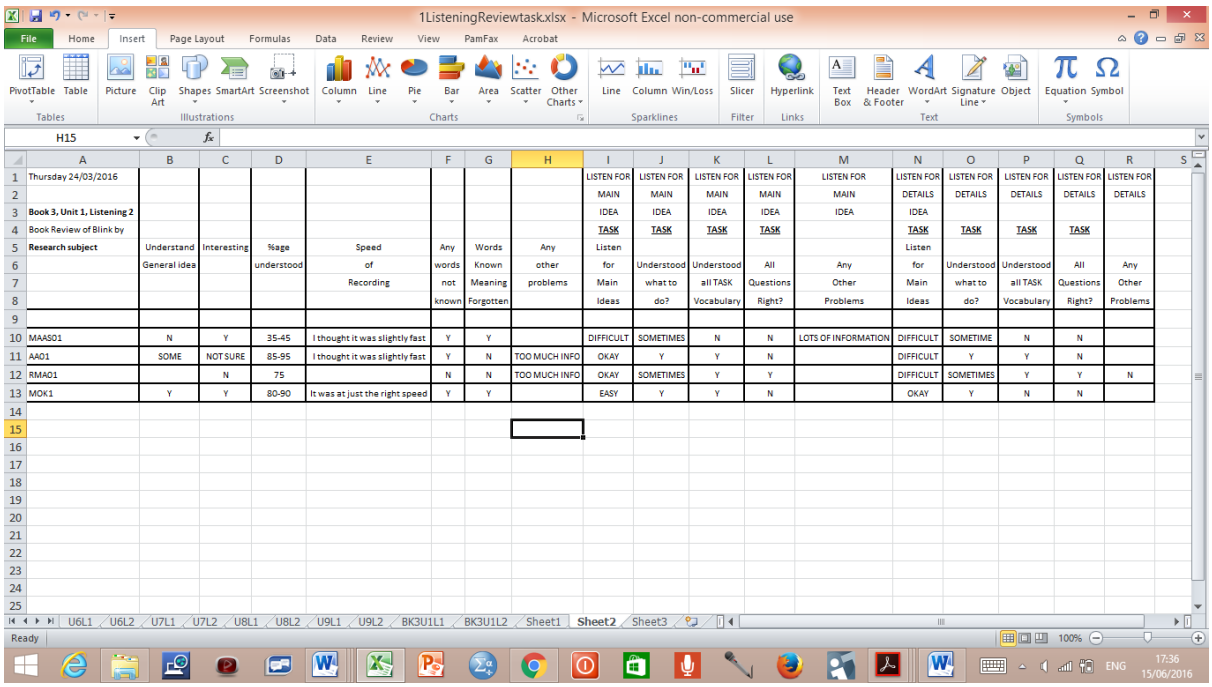
Individual Profile 4 – Main Idea Tasks ABAL02 – CEFR C1

Lesson	MAIN IDEA **TASK** Difficulty	MAIN IDEA **TASK** Understood what to do?	MAIN IDEA **TASK** Understood Vocabulary	MAIN IDEA **TASK** All Questions Right?	MAIN IDEA **TASK** Any Other Problems
BK2U6L1	EASY	N/A	N/A	-	N/A
BK2U6L2	EASY	-	-	-	-
BK2U7L1	-	-	-	-	-
BK2U7L2	EASY	Y	Y	Y 4/4	-
BK2U8L1	EASY	Y	Y	Y	-
BK2U8L2	EASY	Y	Y	Y	-
BK2U9L1	EASY	Y	Y	Y	-
BK2U9L2	EASY	Y	Y	Y	-
BK3U1L1	EASY	Y	Y	Y	-
BK3U1L2	-	-	-	-	-

Individual Profile 4 – Detail Tasks - ABAL02 – CEFR C1

Lesson	LISTEN FOR DETAILS **TASK** Difficulty	LISTEN FOR DETAILS **TASK** Understood what to do?	LISTEN FOR DETAILS **TASK** Understood Vocabulary	LISTEN FOR DETAILS **TASK** All Questions Right?	LISTEN FOR DETAILS **TASK** Any Other Problems
BK2U6L1	EASY	N/A	N/A	N/A	N/A
BK2U6L2	HARD	-	-	-	-
BK2U7L1	-	-	-	-	-
BK2U7L2	OKAY	Y	Y	Y	N
BK2U8L1	EASY	Y	Y	N	-
BK2U8L2	EASY	Y	Y	Y	-
BK2U9L1	EASY	Y	Y	Y	-
BK2U9L2	EASY	Y	Y	Y	-
BK3U1L1	-	Y	Y	Y	-
BK3U1L2	-	-	-	-	-

Appendix N – Sample listening review data



1ListeningReviewtask.xlsx - Microsoft Excel non-commercial use

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Thursday 24/03/2016								LISTEN FOR	LISTEN FOR	LISTEN FOR	LISTEN FOR	LISTEN FOR	LISTEN FOR	LISTEN FOR	LISTEN FOR	LISTEN FOR	LISTEN FOR
2									MAIN	MAIN	MAIN	MAIN	MAIN	DETAILS	DETAILS	DETAILS	DETAILS	DETAILS
3	**Book 3, Unit 1, Listening 2**								IDEA	IDEA	IDEA	IDEA	IDEA	IDEA				
4	Book Review of Blink by								TASK	TASK	TASK	TASK		TASK	TASK	TASK	TASK	
5	**Research subject**	Understand	Interesting	%age	Speed	Any	Words	Any	Listen					Listen				
6		General idea		understood	of	words	Known	other	for	Understood	Understood	All	Any	for	Understood	Understood	All	Any
7					Recording	not	Meaning	problems	Main	what to	all TASK	Questions	Other	Main	what to	all TASK	Questions	Other
8						known	Forgotten		Ideas	do?	Vocabulary	Right?	Problems	Ideas	do?	Vocabulary	Right?	Problems
9																		
10	MAAS01	N	Y	35-45	I thought it was slightly fast	Y	Y		DIFFICULT	SOMETIMES	N	N	LOTS OF INFORMATION	DIFFICULT	SOMETIME	N	N	
11	AA01	SOME	NOT SURE	85-95	I thought it was slightly fast	Y	N	TOO MUCH INFO	OKAY	Y	Y	N		DIFFICULT	Y	Y	N	
12	RMA01		N	75		N	N	TOO MUCH INFO	OKAY	SOMETIMES	Y	Y		DIFFICULT	SOMETIMES	Y	Y	N
13	MOK1	Y	Y	80-90	It was at just the right speed	Y	Y		EASY	Y	Y	N		OKAY	Y	N	N	
14																		
15																		
16																		
17																		
18																		
19																		
20																		
21																		
22																		
23																		
24																		
25																		

Appendix O - CEFR descriptors for the listening skill taken from the OPT

A0 (High) - Working towards A1
A1 – Can typically understand in familiar settings, e.g. daily routine, school, work, family: • Basic words and phrases, • Relationships between speakers, • Speech at significantly slower than native speaker speed
A2 – In addition to competencies from the previous level, can typically understand in mostly familiar settings, e.g. shopping, past events, holidays: • Basic intended meanings • Situations and interpersonal relations • Basic expression of feelings, opinions, advice, and problems • Speech at slower than native speaker speed
B1 – In addition to competencies from the previous levels, can typically understand in familiar and some less familiar settings, e.g. technology, current affairs: • A wider range of intended meanings • Simple meanings implied within or beyond the listening text • Situations and interpersonal relations • Expression of persuasion, warning, reasons, agreement • Basic organisation of the listening text
B2 – In addition to competencies from previous levels, can typically understand in familiar and unfamiliar settings, e.g. customs and traditions, history: • A wide range of intended meanings • A range of meanings implied within or beyond the listening text • A wide range of attitudinal meanings • The purpose and organisation of the text • Speech at native speaker speed
C1 – In addition to competencies from previous levels, can typically understand in complex and some abstract settings: • A very wide range of intended meanings • Complex meanings implied within or beyond the listening text • A wide range of complex attitudinal meanings • Complex organisation of the listening text • Speech at native speaker speed

Appendix P - Listening passage analysis

Listening passage Analysis BK=book U=unit L=listening passage	No. of words	Total Length **(mins)**	Total length **(secs)**	length in seconds (note1)	Start-end times	Words per min	Words per sec	Speed Range words per second & average
BK2 U6 L 1	330	2.46	166	1st min	0.09-1.09	121	2.02	1.90-2.13
				2nd min	1.10-2.09	114	1.90	Avg= 2.00
				3rd min	2.10-2.55	96 (45 secs)	2.13	
BK2 U6 L2	624	4.32	272	1st min	0.9-1.09	137	2.28	2.12-2.58
				2nd min	1.10-2.09	133	2.22	Avg= 2.33
				3rd min	2.10-3.09	127	2.12	
				4th min	3.10-4.09	147	2.45	
				5th min	4.10-4.41	80 (31 secs)	2.58	
BK2 U7 L!	393	2.20	160	1st min	0.11-1.11	122	2.03	2.13-2.38
				2nd min	1.12-2.12	143	2.38	Avg= 2.18
				3rd min	2.13-3.03	128 (50 secs)	2.13	
BK2 U7 L2	403	3.27	207	1st min	0.11-1.11	114	1.9	1.88-2.22
				2nd min	1.12-2.12	116	1.93	Avg= 1.98
				3rd min	2.13-	113	1.88	

					3.13			
				4th min	3.14-3.41	60 (27 secs)	2.22	
BK2 U8 L1	485	3.48	228	1st min	0.09-1.09	124	2.06	1.78-2.25
				2nd min	1.10-2.10	107	1.78	Avg-2.06
				3rd min	2.11-3.11	130	2.17	
				4th min	3.12-4.07	124 (55 secs)	2.25	
Listening passage Analysis BK=book U=unit L=listening passage	No. of words	Total Length **(mins)**	Total length **(secs)**	length in seconds (note1)	Start-end times	Words per min	Words per sec	Speed Range words per second & average
BK2 U8 L2	600	5.04	304	1st min	0.10-1.10	132	2.2	1.78=2.2
				2nd min	1.11-2.11	124	2.00	Avg-1.97
				3rd min	2.12-3.12	115	1.92	
				4th min	3.13-4.13	107	1.78	
				5th min	4.14-5.14	118	1.97	
				6th min	5.15-5.18	4 (3 secs)	-	Exclude
BK2 U9 L1	531	4.03	243	1st min	0.09-1.09	129	2.15	2.15-2.33
				2nd min	1.10-2.10	140	2.33	Avg-2.2
				3rd min	2.11-3.11	130	2.17	
				4th min	3.12-4.12	129	2.15	

				5th min	4.13-4.14	3 (2 sec)	-	Exclude from avg
BK2 U9 L2	450	3.51	231	1st min	0.09-1.09	118	1.97	1.85-2.04
				2nd min	1.10-2.10	111	1.85	Avg 1.95
				3rd min	2.11-3.11	117	1.95	
				4th min	3.12-4.03	104 (45 secs)	2.04	
				5th min				
BK3 U1 L1	524	3.43	223	1st min	0.10-1.10	133	2.22	2.22-2.45
				2nd min	1.11-2.11	146	2.43	Avg 2.34
				3rd min	2.12-3.12	147	2.45	
				4th min	3.13-3.56	98 (44 secs)	2.28	
BK3 U1 L2	479	3.46	226	1st min	0.11-1.11	128	2.13	1.97-2.37
				2nd min	1.12-2.12	124	2.07	Avg 2.13
				3rd min	2.13-3.13	118	1.97	
				4th min	3.14-4.00	109 (46 secs)	2.37	

Appendix Q - Ranking of listening passages by average words per minute (Ascending)

Book No.	Chp Unit	Pass no.	Listening Passage name	Learners present (out of 9)	Words rate/sec Avg	Fastest speech rate/sec	Slowest speech rate/sec	Length minutes	Length words
2	9	2	Can numbers lie	4	1.95	2.04	1.85	3.51	450
2	8	2	The Great Banana Race	4	1.97	2.2	1.78	5.04	600
2	7	2	The Great Pacific Garbage Patch	9	1.98	2.22	1.88	3.27	403
2	6	1	Howtoons	8	2	2.13	1.9	2.46	330
2	8	1	The Art of Storytelling	5	2.06	2.25	1.78	3.48	485
3	1	2	Book Review of Blink by Malcolm Gladwell	6	2.13	2.37	1.97	3.46	479
2	7	1	Sustainable Dave	7	2.18	2.38	2.13	2.2	393
2	9	1	Personal numbers	4	2.22	2.33	2.15	4.03	531
2	6	2	Sell-it-yourself	2	2.33	2.58	1.12	4.32	624
3	1	1	The Psychology of first impressions	4	2.34	2.45	2.22	3.43	524

Appendix R - Listening passages and task types

Listening text Analysis	Name	Listening text type	Number of speakers	Task type Main Ideas	Task type Details
BK2 U6 L1	Howtoons	Lecture /informational	1	Select 3 main ideas from 6	Cloze questions
BK2 U6 L2	Sell-it-yourself	Interview	2	Select 1 from 3 paragraphs that expresses main idea	Write short notes based on prompts
BK2 U7 L1	Sustainable Dave	Conversation	2	MCQ's	T/F
BK2 U7 L2	The Great Pacific Garbage Patch	Lecture/Informational	1	MCQ's – choose from a,b,c	Listen complete sentence – MCQ (2 responses)
BK2 U8 L1	The Art of storytelling	Lecture by professor	1	Select 3 sentences that best express the main ideas from 6	MCQs
BK2 U8 L1	The Great Banana Race	Radio show	2	T/F	MCQs
BK2 U9 L1	Personal numbers	Radio show / interview	7	MCQs	Match numbers to explanation of numbers
BK2 U9 L2	Can numbers lie?	Lecture by professor	1	T/F	MCQs
BK3 U1 L1	The psychology of first impressions	Lecture / informational	1	T/F	MCQs
BK3 U1 L2	Book review of Blink by Malcolm Gladwell	Radio show	2	MCQ	Match detail to example

Appendix T - Profile of participants for each passage by CEFR listening band

Passage	A0(High)	A1	A2	B1	C1	N/A	Total
BK2U6L1	2	1	1	2	1	1	8
BK2U6L2	2	1	1	3	1	1	9
BK2U7L1	2	1	1	3	0	1	7
BK2U7L2	2	1	1	3	1	1	9
BK2U8L1	2	1	1	3	1	0	8
BK2U8L2	2	1	1	1	1	1	7
BK2U9L1	2	1	1	1	1	1	7
BK2U9L2	2	1	1	1	1	1	7
BK3U1L1	2	0	0	2	1	1	6
BK3U1L2	2	0	1	3	0	1	7
Total	19	8	9	22	8	9	75

Appendix U - Phoneme Test word list

1	Pace
2	Brace
3	Pack
4	Back
5	Pad
6	Bad
7	Page
8	Badge
9	Pain
10	Pair
11	Bear
12	Pale
13	Bale
14	Palm
15	Balm
16	Pan
17	Ban
18	People
19	Rob
20	Lob
21	Park
22	Bark
23	Balance
24	Palance
25	Big
26	Pig
27	Bang
28	Pang

Appendix V – Phoneme Results

Research Participant Code	Phoneme Test Score ?/28
MOAL01	N/A
ABAL01	26
SAAL01	20
YUGE01	26
OMAL01	16
ABAL02	17
AHMO01	21
RAAL01	19
MOKH01	21

Appendix W – Passages ranked by speech rate – Slowest first

Book No.	Chp Unit	Pass no.	Listening Passage name	Learners present (out of 9)	Words rate/sec Avg	Fastest speech rate/sec	Slowest speech rate/sec	Length minutes	Length words
2	9	2	Can numbers lie	4	1.95	2.04	1.85	3.51	450
2	8	2	The Great Banana Race	4	1.97	2.2	1.78	5.04	600
2	7	2	The Great Pacific Garbage Patch	9	1.98	2.22	1.88	3.27	403
2	6	1	Howtoons	8	2	2.13	1.9	2.46	330
2	8	1	The Art of Storytelling	5	2.06	2.25	1.78	3.48	485
3	1	2	Book Review of Blink by Malcolm Gladwell	6	2.13	2.37	1.97	3.46	479
2	7	1	Sustainable Dave	7	2.18	2.38	2.13	2.2	393
2	9	1	Personal numbers	4	2.22	2.33	2.15	4.03	531
2	6	2	Sell-it-yourself	2	2.33	2.58	1.12	4.32	624
3	1	1	The Psychology of first impressions	4	2.34	2.45	2.22	3.43	524

Appendix X – Listening passage lexical measures – Vocab Profiler

Vocab Pofiler							**Lex density**
	Overall % K1	Overall % K2	Overall % AWL Acad	Overall % Off-list	Words in text Tokens	Type-token ratio	**Content/ words**
BK1U6L1	75.83	6.95	4.53	12.69	331	0.56	0.56
BK2U6L2	84.51	3.37	5.21	6.9	652	0.4	0.49
BK2U7L1	77.57	5.49	1.91	15.04	419	0.47	0.58
BK2U7L2	78.47	5.69	1.73	14.11	404	0.48	0.57
BK2U8L1	91.16	4.22	0.4	4.22	498	0.38	0.44
BK2U8L2	82.68	4.41	1.31	11.6	612	0.4	0.51
BK2U9L1	91.13	4.91	0.94	3.02	530	0.41	0.49
BK2U9L2	84.53	9.59	2.61	3.27	459	0.44	0.54
BK3U1L1	86.47	6.95	3.2	3.38	532	0.37	0.46
BK3U1L2	83.85	4.55	3.11	8.49	483	0.47	0.54

Appendix Y – Listening passage lexical measures - Cohmetrix

Cohmetrix				
	MTLD	VOCD	Word Concreteness	Word Concreteness
	LDMTLD 'Lexical diversity, MTLD, all words'	LDVOCD 'Lexical diversity, VOCD, all words'	z score	%age
			PCCNCz 'Text Easability PC Word concreteness, z score'	PCCNCp 'Text Easability PC Word concreteness, percentile'
BK1U6L1	93.873	109.455	0.695	75.49
BK2U6L2	66.763	103.696	-0.612	27.09
BK2U7L1	75.871	88.258	0.324	62.55
BK2U7L2	49.522	74.287	0.911	81.86
BK2U8L1	45.818	66.293	0.164	56.36
BK2U8L2	63.171	95.513	0.358	63.68
BK2U9L1	61.03	86.857	-0.173	43.25
BK2U9L2	87.588	124.583	0.108	53.98
BK3U1L1	67.756	85.768	-0.621	26.76
BK3U1L2	81.858	124.42	-0.637	26.43

Appendix Z – Listening passage lexical measures – Text Inspector

Text Inspector	Syllable count (lexical)	No. of seconds (manual)	Syllables per second (Manual calculation)	Words with more than 2 syllables - Percentage	Average syllables per sentence
BK1U6L1	493	166	2.97	9.67	18.96
BK2U6L2	883	272	3.25	8.73	14.24
BK2U7L1	528	160	3.30	5.24	11.48
BK2U7L2	578	207	2.79	8.13	15.62
BK2U8L1	680	228	2.98	7.14	17
BK2U8L2	843	304	2.77	10.89	12.77
BK2U9L1	714	243	2.94	5.33	13.22
BK2U9L2	636	231	2.75	6.71	16.31
BK3U1L1	732	223	3.28	7.68	19.26
BK3U1L2	728	226	3.22	11.68	18.67

Appendix AA – Percentage understood by learner. Comparison across listening passages

%age Understood by Learners

	CEFR	BK2U6L1	BK2U6L2	BK2U7L1	BK2U7L2	BK2U8L1	BK2U8L2	BK2U9L1	BK2U9L2	BK3U1L1	BK3U1L2
MOAL01	N/A	50	25	50	75	-	75	75	50	50	35-45
SAAL01	A0 High	35-45	0	-	25	50	25	75	50	25	25
YUGE01	A0 High	25	10-15	25-35%	25-35	30-40	30	40	30	40	30-35
OMAL01	A1	25	0	25	25	25	25	25	25	-	-
ABAL01	B1	-	100	-	-	100	100	-	-	100	75
ABAL02	C1	85-95	85-95	-	80-90	90-100	90-100	99-100	100	85-95	-
AHMO01	B1	100	100	100	100	100	-	100	100	100	75
RAAL01	B1	100	100	100	100	100	-	-	-	-	75
MOKH01	A2	60-75	80-90	100	95	85-95	75	50	50	-	80-90

Appendix AB – Interesting Topic - descriptive and inferential statistics

Interesting Topic

Passage	Mean	Std. Deviation	Minimum	Maximum	Std. Error of Mean	Range	Variance	N
BK2U6L1	3.00	.926	Not Sure 1	Yes 4	.327	3	.857	8
BK2U6L2	2.75	1.035	Not Sure 1	Yes 4	.366	3	1.071	8
BK2U7L1	3.43	1.134	Not Sure 1	Yes 4	.429	3	1.286	7
BK2U7L2	3.71	.488	Some 3	Yes 4	.184	1	.238	7
BK2U8L1	3.57	.787	No 2	Yes 4	.297	2	.619	7
BK2U8L2	3.00	1.000	No 2	Yes 4	.577	2	1.000	3
BK2U9L1	3.25	.957	No 2	Yes 4	.479	2	.917	4
BK2U9L2	3.25	1.500	Not Sure 1	Yes 4	.750	3	2.250	4
BK3U1L1	3.40	.548	Some 3	Yes 4	.245	1	.300	5
BK3U1L2	2.20	1.304	Not Sure 1	Yes 4	.583	3	1.700	5
Total	3.17	.994	Not Sure 1	Yes 4	.130	3	.987	58

Appendix AC – Understood General Idea - descriptive and inferential statistics

Understood General Idea

Passage	Mean	Std. Deviation	Minimum	Maximum	Std. Error of Mean	Range	Variance	N
BK2U6L1	1.86	.378	No	Yes	.143	1	.143	7
BK2U6L2	1.63	.518	No	Yes	.183	1	.268	8
BK2U7L1	2.00	.000	Yes	Yes	.000	0	.000	4
BK2U7L2	1.71	.488	No	Yes	.184	1	.238	7
BK2U8L1	1.86	.378	No	Yes	.143	1	.143	7
BK2U8L2	1.80	.447	No	Yes	.200	1	.200	5
BK2U9L1	2.00	.000	Yes	Yes	.000	0	.000	4
BK2U9L2	1.50	.577	No	Yes	.289	1	.333	4
BK3U1L1	1.83	.408	No	Yes	.167	1	.167	6
BK3U1L2	1.40	.548	No	Yes	.245	1	.300	5
Total	1.75	.434	No	Yes	.058	1	.189	57

Appendix AD - %age Understood – descriptive and inferential statistics

Percentage Understood

Passage	Mean	Std. Deviation	Minimum	Maximum	Std. Error of Mean	Range	Variance	N
BK2U6L1	62.500	31.8479	25.0	100.0	11.2599	75.0	1014.286	8
BK2U6L2	57.222	45.6968	.0	100.0	15.2323	100.0	2088.194	9
BK2U7L1	72.143	35.5735	25.0	100.0	13.4455	75.0	1265.476	7
BK2U7L2	70.556	33.9526	25.0	100.0	11.3175	75.0	1152.778	9
BK2U8L1	74.375	32.1200	25.0	100.0	11.3561	75.0	1031.696	8
BK2U8L2	60.714	33.2200	25.0	100.0	12.5560	75.0	1103.571	7
BK2U9L1	66.286	28.9235	25.0	100.0	10.9321	75.0	836.571	7
BK2U9L2	57.857	30.5310	25.0	100.0	11.5396	75.0	932.143	7
BK3U1L1	67.500	33.1285	25.0	100.0	13.5247	75.0	1097.500	6
BK3U1L2	58.571	24.2752	25.0	85.0	9.1752	60.0	589.286	7
Total	64.787	32.2265	.0	100.0	3.7212	100.0	1038.548	75

Appendix AE - Speed of Recording (Speech rate) – descriptive and inferential statistics

Speed Of Recording (Speech Rate)

Speed

Passage	Mean	Std. Deviation	Minimum	Maximum	Std. Error of Mean	Range	Variance	N
BK2U6L1	3.50	.756	Just right (Score 3)	Very fast (Score 5)	.267	2	.571	8
BK2U6L2	3.33	.500	Just right (Score 3)	Slightly fast (Score 4)	.167	1	.250	9
BK2U7L1	3.33	.516	Just right (Score 3)	Slightly fast (Score 4)	.211	1	.267	6
BK2U7L2	3.22	.441	Just right (Score 3)	Slightly fast (Score 4)	.147	1	.194	9
BK2U8L1	3.13	.354	Just right (Score 3)	Slightly fast (Score 4)	.125	1	.125	8
BK2U8L2	3.17	.408	Just right (Score 3)	Slightly fast (Score 4)	.167	1	.167	6
BK2U9L1	3.00	.000	Just right (Score 3)	Just right (Score 3)	.000	0	.000	7
BK2U9L2	3.14	.378	Just right (Score 3)	Slightly fast (Score 4)	.143	1	.143	7
BK3U1L1	3.17	.408	Just right (Score 3)	Slightly fast (Score 4)	.167	1	.167	6
BK3U1L2	3.50	.837	Just right (Score 3)	Very fast (Score 5)	.342	2	.700	6
Total	3.25	.496	Just right (Score 3)	Very fast (Score 5)	.059	2	.246	72

Appendix AF - Words not known (new words) – descriptive and inferential statistics

Words Not Known

Passage	Mean	Std. Deviation	Minimum	Maximum	Std. Error of Mean	Range	Variance	N
BK2U6L1	1.750	.4629	All words known	New Words	.1637	1.0	.214	8
BK2U6L2	2.000	.0000	New Words	New Words	.0000	.0	.000	9
BK2U7L1	1.429	.5345	All words known	New Words	.2020	1.0	.286	7
BK2U7L2	1.889	.3333	All words known	New Words	.1111	1.0	.111	9
BK2U8L1	1.875	.3536	All words known	New Words	.1250	1.0	.125	8
BK2U8L2	1.857	.3780	All words known	New Words	.1429	1.0	.143	7
BK2U9L1	1.857	.3780	All words known	New Words	.1429	1.0	.143	7
BK2U9L2	1.857	.3780	All words known	New Words	.1429	1.0	.143	7
BK3U1L1	1.833	.4082	All words known	New Words	.1667	1.0	.167	6
BK3U1L2	1.857	.3780	All words known	New Words	.1429	1.0	.143	7
Total	1.827	.3811	All words known	New Words	.0440	1.0	.145	75

Appendix AG - Words known but meanings forgotten – descriptive and inferential statistics

Words Known Meanings Forgotten

Passage	Mean	Std. Deviation	Minimum	Maximum	Std. Error of Mean	Range	Variance	N
BK2U6L1	1.625	.5175	Vocab Meanings NOT Forgotten	Vocab Meanings Forgotten	.1830	1.0	.268	8
BK2U6L2	1.778	.4410	Vocab Meanings NOT Forgotten	Vocab Meanings Forgotten	.1470	1.0	.194	9
BK2U7L1	1.000	.0000	Vocab Meanings NOT Forgotten	Vocab Meanings NOT Forgotten	.0000	.0	.000	7
BK2U7L2	1.556	.5270	Vocab Meanings NOT Forgotten	Vocab Meanings Forgotten	.1757	1.0	.278	9
BK2U8L1	1.750	.4629	Vocab Meanings NOT Forgotten	Vocab Meanings Forgotten	.1637	1.0	.214	8
BK2U8L2	1.714	.4880	Vocab Meanings NOT Forgotten	Vocab Meanings Forgotten	.1844	1.0	.238	7
BK2U9L1	1.714	.4880	Vocab Meanings NOT Forgotten	Vocab Meanings Forgotten	.1844	1.0	.238	7
BK2U9L2	1.857	.3780	Vocab Meanings NOT Forgotten	Vocab Meanings Forgotten	.1429	1.0	.143	7
BK3U1L1	1.833	.4082	Vocab Meanings NOT Forgotten	Vocab Meanings Forgotten	.1667	1.0	.167	6
BK3U1L2	1.429	.5345	Vocab Meanings NOT Forgotten	Vocab Meanings Forgotten	.2020	1.0	.286	7
Total	1.627	.4869	Vocab Meanings NOT Forgotten	Vocab Meanings Forgotten	.0562	1.0	.237	75

Appendix AH – Spearman's rho correlation – Listening perception measures

Spearman's Rho Correlations

			InterestingTopic	Understood general idea	%age understood	Speed	Words not known	Meaning forgotten
Spearman's rho	InterestingTopic	Correlation Coefficient	1.000	.318*	.488**	-.187	-.211	-.112
		Sig. (2-tailed)	.	.028	.000	.172	.112	.404
		N	58	48	58	55	58	58
	Understood general idea	Correlation Coefficient	.318*	1.000	.675**	-.204	-.263*	-.351**
		Sig. (2-tailed)	.028	.	.000	.132	.048	.007
		N	48	57	57	56	57	57
	%age understood	Correlation Coefficient	.488**	.675**	1.000	-.358**	-.453**	-.561**
		Sig. (2-tailed)	.000	.000	.	.002	.000	.000
		N	58	57	75	72	75	75
	Speed	Correlation Coefficient	-.187	-.204	-.358**	1.000	.151	.116
		Sig. (2-tailed)	.172	.132	.002	.	.204	.334
		N	55	56	72	72	72	72
	Words not known	Correlation Coefficient	-.211	-.263*	-.453**	.151	1.000	.448**
		Sig. (2-tailed)	.112	.048	.000	.204	.	.000
		N	58	57	75	72	75	75
	Meaning forgotten	Correlation Coefficient	-.112	-.351**	-.561**	.116	.448**	1.000

	Sig. (2-tailed)	.404	.007	.000	.334	.000	.
	N	58	57	75	72	75	75

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix AI – Spearman's Rank correlation - CEFR Correlations

Spearman's Rho Correlations

			CEFRNUMBER
	CEFRNUMBER	Correlation Coefficient	1.000
		Sig. (2-tailed)	
		N	75
	Understood general idea	Correlation Coefficient	.495**
		Sig. (2-tailed)	.000
		N	57
	Interesting Topic	Correlation Coefficient	.355**
		Sig. (2-tailed)	.006
		N	58
	%age understood	Correlation Coefficient	.774**
		Sig. (2-tailed)	.000
		N	75
	Speed	Correlation Coefficient	-.435**
		Sig. (2-tailed)	.000
		N	72
	Words not known	Correlation Coefficient	-.397**
		Sig. (2-tailed)	.000
		N	75
	Meaning forgotten	Correlation Coefficient	-.496**
		Sig. (2-tailed)	.000
		N	75
	Main Idea Questions	Correlation Coefficient	.714**
		Sig. (2-tailed)	.000
		N	68
	MI UNDERSTOODTASK	Correlation Coefficient	.629**
		Sig. (2-tailed)	.000
		N	50
	MI UNDERSTOOD TASKVOCAB	Correlation Coefficient	.759**
		Sig. (2-tailed)	.000
		N	41
	MI ALLQUESTIONS	Correlation Coefficient	.474**
	RIGHT?	Sig. (2-tailed)	.000
		N	50
	Listen	Correlation Coefficient	.523**
	For	Sig. (2-tailed)	.000
	Detail Questions	N	66
	D UNDERSTOOD TASK	Correlation Coefficient	.620**
		Sig. (2-tailed)	.000
		N	49
	D UNDERSTOOD TASK VOCAB	Correlation Coefficient	.748**
		Sig. (2-tailed)	.000

	N	43
D ALL QUESTIONS RIGHT?	Correlation Coefficient	.293*
	Sig. (2-tailed)	.035
	N	52

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Appendix AJ – Spearman's Rank Correlation Coefficient – Lexical measures

		Ranking Understood Gen Idea	Rank Percent Understood	Rank New Words Not Known	Rank Speed
Spearmans rho	Correlation Coefficient	1.000	1.000	1.000	1.000
	Sig. (2-tailed)				
	N	10	10	10	10
K1RANK	Correlation Coefficient	-.320	-.030	-.013	.079
	Sig. (2-tailed)	.367	.934	.973	.828
	N	10	10	10	10
K2RANK	Correlation Coefficient	-.209	.042	**.557**	-.018
	Sig. (2-tailed)	.562	.907	.095	.960
	N	10	10	10	10
AWLRANK	Correlation Coefficient	-.160	**-.576**	.156	**-.671***
	Sig. (2-tailed)	.659	.082	.666	.034
	N	10	10	10	10
OFFLISTRANK	Correlation Coefficient	.012	.261	.206	**-.720***
	Sig. (2-tailed)	.973	.467	.567	.019
	N	10	10	10	10
TYPE	Correlation Coefficient	-.121	-.049	.239	**-.598**
TOKEN	Sig. (2-tailed)	.740	.894	.506	.068
RANK	N	10	10	10	10
MTLD	Correlation Coefficient	-.215	**-.491**	**.619**	**-.567**
RANK	Sig. (2-tailed)	.550	.150	.056	.087
	N	10	10	10	10
VOCD_RANK	Correlation Coefficient	-.382	**-.830****	.213	**-.476**
	Sig. (2-tailed)	.277	.003	.555	.165
	N	10	10	10	10
Word	Correlation Coefficient	.135	**.382**	.056	-.079
Concrete	Sig. (2-tailed)	.709	.276	.877	.828
Percent_RANK	N	10	10	10	10
2SyllPerc_RANK	Correlation Coefficient	**-.566**	**-.479**	-.219	**-.591**
	Sig. (2-tailed)	.088	.162	.544	.072
	N	10	10	10	10

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix AK – Rankings of variables

Passage	Passage Code	Overall % K1	K1RANK
1	BK2U6L1	75.83	10
2	BK2U6L2	84.51	5
3	BK2U7L1	77.57	9
4	BK2U7L2	78.47	8
5	BK2U8L1	91.16	1
6	BK2U8L2	82.68	7
7	BK2U9L1	91.13	2
8	BK2U9L2	84.53	4
9	BK3U1L1	86.47	3
10	BK3U1L2	83.85	6
Passage	Passage Code	Overall % K2	RANK
1	BK2U6L1	6.95	2
2	BK2U6L2	3.37	10
3	BK2U7L1	5.49	5
4	BK2U7L2	5.69	4
5	BK2U8L1	4.22	9
6	BK2U8L2	4.41	8
7	BK2U9L1	4.91	6
8	BK2U9L2	9.59	1
9	BK3U1L1	6.95	3
10	BK3U1L2	4.55	7
Passage	Passage Code	Overall % AWL Acad	RANK
1	BK2U6L1	4.53	2
2	BK2U6L2	5.21	1
3	BK2U7L1	1.91	6
4	BK2U7L2	1.73	7
5	BK2U8L1	0.4	10
6	BK2U8L2	1.31	8
7	BK2U9L1	0.94	9
8	BK2U9L2	2.61	5
9	BK3U1L1	3.2	3
10	BK3U1L2	3.11	4

Appendix AK – Rankings of variables (continued)

Passage	Passage Code	Overall % Off list	RANK
1	BK2U6L1	12.69	3
2	BK2U6L2	6.9	6
3	BK2U7L1	15.04	1
4	BK2U7L2	14.11	2
5	BK2U8L1	4.22	7
6	BK2U8L2	11.6	4
7	BK2U9L1	3.02	10
8	BK2U9L2	3.27	9
9	BK3U1L1	3.38	8
10	BK3U1L2	8.49	5

Passage	Passage Code	Type-token Ratio	RANK
1	BK2U6L1	0.56	1
2	BK2U6L2	0.4	6
3	BK2U7L1	0.47	3
4	BK2U7L2	0.48	2
5	BK2U8L1	0.38	7
6	BK2U8L2	0.4	6
7	BK2U9L1	0.41	5
8	BK2U9L2	0.44	4
9	BK3U1L1	0.37	8
10	BK3U1L2	0.47	3

Appendix AK – Rankings of variables (continued)

		Word Concreteness % PCCNCp 'Text Easability PC Word concreteness, percentile'	WordConcretePercent_RANK
1	BK2U6L1	75.49	2
2	BK2U6L2	27.09	8
3	BK2U7L1	62.55	4
4	BK2U7L2	81.86	1
5	BK2U8L1	56.36	5
6	BK2U8L2	63.68	3
7	BK2U9L1	43.25	7
8	BK2U9L2	53.98	6
9	BK3U1L1	26.76	9
10	BK3U1L2	26.43	10
Pass. No		Words more than 2 syll %	Rank
1	BK1U6L1	9.67	3
2	BK2U6L2	8.73	4
3	BK2U7L1	5.24	10
4	BK2U7L2	8.13	5
5	BK2U8L1	7.14	7
6	BK2U8L2	10.89	2
7	BK2U9L1	5.33	9
8	BK2U9L2	6.71	8
9	BK3U1L1	7.68	6
10	BK3U1L2	11.68	1

Appendix AK – Rankings of variables (continued)

		MTLD LDMTLD 'Lexical diversity, MTLD, all words'	RANK
1	BK2U6L1	93.873	1
2	BK2U6L2	66.763	6
3	BK2U7L1	75.871	4
4	BK2U7L2	49.522	9
5	BK2U8L1	45.818	10
6	BK2U8L2	63.171	7
7	BK2U9L1	61.03	8
8	BK2U9L2	87.588	2
9	BK3U1L1	67.756	5
10	BK3U1L2	81.858	3
		VOCD LDVOCD 'Lexical diversity, VOCD, all words'	VOCD_RANK
1	BK2U6L1	109.455	3
2	BK2U6L2	103.696	4
3	BK2U7L1	88.258	6
4	BK2U7L2	74.287	9
5	BK2U8L1	66.293	10
6	BK2U8L2	95.513	5
7	BK2U9L1	86.857	7
8	BK2U9L2	124.583	1
9	BK3U1L1	85.768	8
10	BK3U1L2	124.42	2

Appendix AL – Ranking of listening passages by listening perception measures

R=Rank

	Interesting		Understood		%age		Speed		New Words		Meaning	
	Topic		General idea		Understood				not known		Forgotten	
R	Passage	Mean	**R**	Passage	Mean	**R**	Passage	Mean	**R**	Passage	Mean	**R**	Passage	Mean	**R**	Passage	Mean
1	BK2U7L2	3.71	1	BK2U7L1	2	1	BK2U8L1	74.375	1	BK2U9L1	3	1	BK2U7L1	1.429	1	BK2U7L1	1
2	BK2U8L1	3.57	1	BK2U9L1	2	2	BK2U7L1	72.143	2	BK2U8L1	3.13	2	BK2U6L1	1.75	2	BK3U1L2	1.429
3	BK2U7L1	3.43	2	BK2U6L1	1.86	3	BK2U7L2	70.556	3	BK2U9L2	3.14	3	BK3U1L1	1.833	3	BK2U7L2	1.556
4	BK3U1L1	3.4	2	BK2U8L1	1.86	4	BK3U1L1	67.5	4	BK2U8L2	3.17	4	BK2U8L2	1.857	4	BK2U6L1	1.625
5	BK2U9L1	3.25	3	BK3U1L1	1.83	5	BK2U9L1	66.286	4	BK3U1L1	3.17	4	BK2U9L1	1.857	5	BK2U8L2	1.714
5	BK2U9L2	3.25	4	BK2U8L2	1.8	6	BK2U6L1	62.5	5	BK2U7L2	3.22	4	BK2U9L2	1.857	5	BK2U9L1	1.714
6	BK2U6L1	3	5	BK2U7L2	1.71	7	BK2U8L2	60.714	6	BK2U6L2	3.33	4	BK3U1L2	1.857	6	BK2U8L1	1.75
6	BK2U8L2	3	6	BK2U6L2	1.63	8	BK3U1L2	58.571	6	BK2U7L1	3.33	5	BK2U8L1	1.875	7	BK2U6L2	1.778
7	BK2U6L2	2.75	7	BK2U9L2	1.5	9	BK2U9L2	57.857	7	BK2U6L1	3.5	6	BK2U7L2	1.889	8	BK3U1L1	1.833
8	BK3U1L2	2.2	8	BK3U1L2	1.4	10	BK2U6L2	57.222	7	BK3U1L2	3.5	7	BK2U6L2	2	9	BK2U9L2	1.857

1=Not sure	1=No	1=very slow	1=known	1=NOT forgotten
2=No	2=Yes	2=slightly slow	2=not	2=Forgotten
3=Some		3=Just right		
4=Yes		4=Slightly fast		
		5=Very fast		

Appendix AM – Examples of Rating Scale Descriptors – Brindley

Examples of Brindley's rating scale for a particular level are listed below:

Listening Comprehensions

> Can recognise a few intonation patterns (e.g. Yes/no questions)
>
> Little understanding of syntax. Meaning deduced from juxtaposition of words and context. Still responds to isolated words in connected speech.
>
> Can handle very short, simple, ritual social exchanges but rarely able to understand enough to keep conversation going of his/her own accord.

Characteristic Problems

> Has great difficulty coping with subjects other than immediate priorities.
>
> Often fails to understand questions which require other than a short, concrete answer (e.g., *why* or *how* questions).
>
> Similar sounding words/segments often confused, causing misunderstandings.
>
> (Brindley, 1982:1, cited in Richards, 1983)